THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 59, PART 6 1 November 1947 No. 336

THE OPTICAL PROPERTIES OF AXIALLY SYMMETRIC MAGNETIC PRISMS

PART 1: THE STUDY OF RAYS IN A PLANE OF SYMMETRY, AND ITS APPLICATION TO THE DESIGN OF PRISM β-SPECTROSCOPES

BY R. E. SIDAY,
Birkbeck College

MS. received 30 *January* 1947

ABSTRACT. The optical properties of the magnetic field in the neighbourhood of two equal co-axial magnetic poles of opposite polarity used as a magnetic prism are discussed. In this first part attention is confined to rays in the plane of symmetry between the poles over which the radial component of the field vanishes everywhere.

The focusing properties and aberrations are discussed generally, and detailed calculations are given :

(1) by simple analytical methods for the case of zero pole separation (homogeneous field);
(2) by numerical computation for the case of a pole separation equal to the radius of either pole.

Inferences are drawn concerning the use of such fields in β-spectroscopy, and it is shown that under certain conditions image formation by the prisms is optically corrected over a very wide aperture, or under other conditions can be corrected in compound instruments consiting of the prism and one or more magnetic lenses. Thus very high resolution is possible in such instruments. The overall optimum design for a β-spectroscope, however, cannot be made without a consideration of the focusing towards the plane of symmetry of rays confined to that plane. This focusing will be considered in Part 2.

§ 1. INTRODUCTION

THE magnetic prism has increasing applications in nuclear physics. Already in different forms of the mass spectrometer the magnetic prism plays a fundamental part, and in β-spectroscopy suggestions have been made regarding the use of prisms to obtain high resolving powers. Very little information, however, is available regarding the optical properties, particularly the aberrations, of these prisms.

This note will be confined to a consideration of the use as prisms of axially symmetric fields that possess a plane of symmetry in which the radial component of the field vanishes everywhere. In this first part only trajectories in the plane of symmetry will be considered, and the orbits discussed will be of the non-periodic type according to the classification of Coggershall and Muskat (1944), since for the cases considered the strength of the field will decrease as the radius increases.

It might be useful to describe more clearly what is meant by the term ***prism*** in this connexion. As will be discussed in detail, image formation by rays in the plane of symmetry occurs with varying quality according to the distance of the object from the axis of the field, and of course the usual method of semicircular focusing and its more recent modifications employ special cases of this image formation in which the object is immersed in the field. This paper will be concerned more with image formation when the object is so far from the axis that the field is negligibly strong, although a slight extension will be made in § 6 to the case in which the object is located where the field is weak but not negligible. There is no sharp distinction between semicircular focusing and the cases discussed here, although for the latter the general picture will be of rays entering the prism from a considerable distance from the axis, being deviated and finally emerging from a *virtual* focus. The focus will be virtual because the real image, in general, will occur where the field is not negligible and the rays will be further deviated in their passage away from the axis towards field-free space where they will belong to a virtual image.

In optics, the limit to resolution is imposed entirely by the finite wavelength of light. In the spectroscopy of charged particles, the limit arises entirely from the aberrations of image formation. It is clear that if the prism is to play a fundamental rôle in producing high resolution in any spectrometer, the aberrations of its image formation become of crucial importance.

Most mass spectrometers employ a sector-shaped prism of uniform field strength, the focusing properties of which were established by Herzog (1934). Aston, however, virtually uses a homogeneous axially symmetric field, and the optical properties of this for the particular case of this spectrometer were considered by Sawyer (1936). Coggershall and Muskat have examined ion paths in a wide variety of fields and have proposed new designs of mass spectrograph. In β-spectroscopy, however, the focusing properties of beams of much wider angle are important. Their properties will here be considered generally, and detailed calculations will be given for the two special cases (1) of a homogeneous field, (2) of the field in the neighbourhood of two perfect equal and opposite pole coaxial circular magnetic poles separated by a distance equal to the radius of either.

§ 2. GENERAL FOCUSING PROPERTIES IN THE PLANE OF SYMMETRY

Consider an electron moving in the plane of symmetry. Where the field is negligible at great distance from the axis, the trajectory is a straight line. Thus an electron entering the field of the prism moves initially in a straight line, is curved in the magnetic field, and finally emerges again along another straight trajectory. Because of the axial symmetry, the initial straight trajectory can be specified by a single parameter, *b*, which is the shortest distance of this straight line produced from the axis.

The differential equations of the motion of the electron of specific charge e/m moving in an axially symmetric field are

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{e}{m} r H_z \frac{d\theta}{dt}, \qquad \text{......(1)}$$

$$\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = -\frac{e}{m} r H_r \frac{dz}{dt} + \frac{e}{m} r H_z \frac{dr}{dt}, \qquad \text{......(2)}$$

$$\frac{d^2z}{dt^2}=\frac{\mathbf{e}}{m}rH_r\frac{d\theta}{dt}, \qquad \ldots\ldots(3)$$

where r, θ and z, are cylindrical coordinates defining the position of the particle with respect to the axis of the field and H_r and H_z are respectively the radial and axial components of the field at that point. The equations for a ray in the plane of symmetry where $H_r=0$, reduce to

$$\frac{d^2r}{dt^2}-r\left(\frac{d\theta}{dt}\right)^2=-\frac{\mathbf{e}}{m}rH_z\frac{d\theta}{dt}, \qquad \ldots\ldots(4)$$

$$\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right)=\frac{\mathbf{e}}{m}rH_z\frac{dr}{dt}. \qquad \ldots\ldots(5)$$

Integration of the last equation gives

$$\left|r^2\frac{d\theta}{dt}\right|_1^2=\frac{\mathbf{e}}{m}\int_1^2 rH_z\,dr. \qquad \ldots\ldots(6)$$

If ψ is the angle between the trajectory at any point and the radius vector at that point, and if v is the constant velocity of the electron in the magnetic field

$$r^2\frac{d\theta}{dt}=rv\sin\psi,$$

so that

$$r_1\sin\psi_1-r_2\sin\psi_2=\frac{1}{(H\rho)}\int_2^1 rH_z\,dr \qquad \ldots\ldots(7)$$

where $(H\rho)$ is the "momentum" $mv/\mathbf{e}$. For very large values of r,

$$r_1.\sin\psi=b.$$

Thus equation (7) becomes

$$r\sin\psi-b=\frac{1}{(H\rho)}\int_\infty^r rH_z\,dr, \qquad \ldots\ldots(8)$$

and this equation will subsequently be made the basis of ray tracing. During its passage through the field, the electron will follow a curved path which will at one point have a closest distance from the axis $r_{\min}$ whose value is obtained by putting $\psi=\pi/2$ so that

$$r_{\min}-b=\frac{1}{(H\rho)}\int_\infty^{r\min} rH_z dr.$$

As H_z is a function of r only, the complete trajectory is symmetrical about this point of closest approach to the axis. In particular, the parameter b for the straight line along which the electron moves after it has left the field, has the same value as that of the straight trajectory before it entered the field of the prism. This can be formally shown, for if b and b' are the two values of these parameters respectively

$$b'-r_{\min}=-\frac{1}{(H\rho)}\int_\infty^{r\min} rH_z dr,$$

$$r_{\min}-b=\frac{1}{(H\rho)}\int_\infty^{r\min} rH_z dr,$$

so that $b=b'$.

Thus in a given field any trajectory is defined completely by the two parameters b and $(H\rho)$. The approaching electron of parameter b will be deviated through an angle $2 . \phi_b$, which is equal to $\pi - 2\theta_b$ where θ_b is the total increase in the coordinate θ, as the electron approaches from infinity to the point of closest approach. These angles are marked in figure 1. In general, for a given $(H\rho)$, ϕ_b will depend on b, for which the following sign convention is adopted. If the length of the trajectory in the field is shorter than that of the trajectory directed initially towards the axis (i.e. with $b=0$), then b is taken as positive. If it is longer, b is negative. If $2 . \phi_0$ is the deviation of a principal ray, that is, one initially directed towards the axis, and $2\phi_b$ that of a parallel ray of parameter b, after deviation, these two rays will appear to have come from a focus at a distance equal to

$$f_b = \frac{-b}{\sin(2\phi_b - 2\phi_0)} \qquad \ldots\ldots(9)$$

from the axis. If as an optical convention distances are measured from the origin O, and object distances are positive in a direction opposed to that of the motion of the electron, while image distances are positive in the direction of motion, then if f_b is to be positive for a convergent lens action, the sign of (9) is correct. If now f_b is constant for all values of b, the focus is free from aberration, but in any case the values of ϕ_b as a function of b give all the information required to examine the quality of the actual focus.

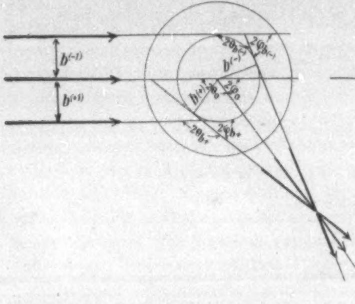


Figure 1. Focusing a parallel bundle.

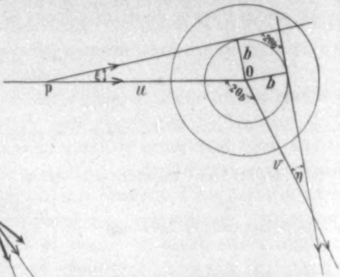


Figure 2. Imaging a point source.

In order to obtain the rules for the focusing of a point not at infinity, consider the two rays of figure 2 coming from the point P. Then $\sin \xi = b/u$, and $\sin \eta = b/v$, where u and v are object and image distances from the axis respectively.

$$\eta + 2\theta_0 + \xi = 2\theta_b,$$

$$\eta + \xi = 2\theta_b - 2\theta_0 = 2\phi_0 - 2\phi_b,$$

$$-\cos\eta \sin\xi - \sin\eta \cos\xi = \sin(2\phi_b - 2\phi_0),$$

$$\frac{\cos\eta}{u} + \frac{\cos\xi}{v} = \frac{1}{f_b}. \qquad \ldots\ldots(10)$$

The aperture defect shows up now in the presence of the cosine terms as well as in the general dependence of f_b on b. If as an approximation $\cos\eta$ is written as $1-\frac{1}{2}b^2/v^2$, and $\cos\xi$ as $1-\frac{1}{2}b^2/u^2$, then, to this order of accuracy, equation (10) becomes

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f_b\left(1-\frac{1}{2}\frac{b^2}{uv}\right)}. \qquad \text{......(11)}$$

If $f_b\cdot(1-\frac{1}{2}b^2/uv)$ is constant and equal to f_0 say, the image is free from aberration of this order. In general, both the terms f_b and $(1-\frac{1}{2}b^2/uv)$ depend on b, and the possibility arises that for a certain value of u the two terms could cooperate to give corrected or partly corrected image formation. The variation of f_b with b will turn out in general to be too large to be essentially affected by the $(1-\frac{1}{2}b^2/uv)$ term, but it will be shown that this is not always so, and that correction can in fact occur. For a given value of u, the conjugate value of v depends according to equation (11) on b, but in the correction term it will be sufficient to take for uv that derived from the approximate equation $\frac{1}{u}+\frac{1}{v}=\frac{1}{f_0}$. Then, as in elementary optics, the product uv is given by $f_0^2(1+m)(1+1/m)$ where m is the magnification, and the maximum effect of the term $(1-\frac{1}{2}b^2uv)$ arises when uv is smallest, that is, when $m=1$ and $u=v$.

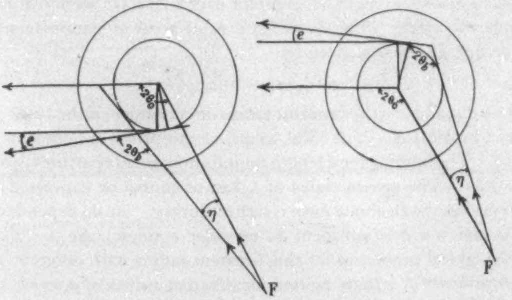


Figure 3. Sign of the aberration.

The condition for the absence of aberration to this order of accuracy can be expressed more conveniently as follows. If u is kept constant, and v_0 and v are the image distances for rays of parameters 0 and b respectively, then from (11)

$$\frac{1}{u}+\frac{1}{v_0}=\frac{1}{f_0},$$

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f_b}\left(1+\frac{1}{2}\frac{b^2}{uv}\right),$$

$$\Delta\left(\frac{1}{v}\right)=\frac{1}{v_0}-\frac{1}{v}=\frac{1}{f_0}-\frac{1}{f_b}-\frac{1}{2}\frac{b^2}{uvf_0},$$

$$\therefore -\Delta v=\frac{v^2}{f_0f_b}\left(\overline{f_b-f_0}-\frac{1}{2}\frac{b^2f_0}{uv}\right). \qquad \text{......(12)}$$

Therefore for corrected image formation it is necessary that

$$f_b - f_0 = \tfrac{1}{2}\frac{b^2 f_0}{uv}, \qquad \ldots\ldots(13)$$

an equation that implies a quadratic form as the functional variation of f_b with b, as well as the correct magnitude of uv. Thus, in general, only partial correction can be expected, although in certain cases full correction does occur.

It will be useful also to consider sometimes the angular aberration. Figure 3 has been drawn in order to fix the magnitude and sign of the angular aberration in a parallel bundle. If this angle is e, then according to the sign of b, e is given by

$$\left.\begin{aligned} e &= \eta - (2\theta_b - 2\theta_0) \quad (b \text{ positive}), \\ e &= \eta - (2\theta_0 - 2\theta_b) \quad (b \text{ negative}), \end{aligned}\right\} \qquad \ldots\ldots(14)$$

and on this convention a positive value of e means that the ray of parameter $\pm b$ *diverges* from the principal ray. Of course the angular aberration in the imaging of a point object at a distance u is simply related to e through the equation

$$\Delta\theta = e - \tfrac{1}{2}\frac{b^3}{f_0 uv}. \qquad \ldots\ldots(15)$$

§ 3. PROPERTIES OF THE HOMOGENEOUS PRISM

Consider a field of strength, H, constant over a circular section of radius a and zero outside this circle. The deviation $2.\phi_b$ of a ray of parameters b and $(H\rho)$ passing through this field is given by

$$\cot\phi_b = (\rho + b)/a\cos\lambda, \qquad \ldots\ldots(16)$$

where $\sin\lambda = b/a$, and ρ is the constant radius of curvature of the track in the field, and is given by $(H\rho)/H$. The focal length of the prism is found by combining (9) and (16). The limiting focal length for infinitely small apertures $\pm b$ is given by $f_0 = (a^2 + \rho^2)/2a$. The general value of f_b can of course be expressed from these two equations but the algebraic form is rather clumsy. As ϕ_b depends only on the rations b/a and ρ/a it is sufficient to consider a single case $a = 10$ cm. and to describe the optical properties for this constant radius with different values of ρ. In figure 4 values of $f_b - f_0$ are plotted for different values of ρ and b, thus giving directly the longitudinal focusing in an initially parallel bundle. For the particular case of $\rho = a$, $\cot\phi_0 = 1$, so that the principal ray is deviated through an angle $\pi/2$. Further, $\cot\phi_b = \sqrt{(a-b)/(a+b)}$, so that f_b becomes constant for all values of $b\,(-a \leqslant b \leqslant +a)$ and takes the value a in agreement with the general expressing for f_0 given above. Thus the aberration completely dissappears over the whole aperture, a result that Korsunsky (1945) has used as the basis of a proposed β-spectroscope.

It is seen from the graphs first, that $f_b - f_0$ steadily increases at first with $\rho - a$, and secondly that the sign of $(f_b - f_0)$ for a given value of ρ changes with the sign of b. The position of minimum confusion lies, therefore, close to the paraxial focus and the spread of the image in this position is terminated sharply by the principal ray. This sharp edge is similar to that associated with images obtained by the method of semicircular focusing. As the sign of the aberration changes with that of $(\rho - a)$, this sharp edge lies on the side of the image spread that is closest to the principal ray belonging to the value of ρ equal to a.

The graphs show that for values of ρ that are not too small, $f_b - f_0$ is roughly proportional to b, so that image formation of a point not at infinity can, according to (13), only be partly corrected, and that for only one side of the aperture. For values of $(\rho - a)/a$ between 0 and 0·05 approximately, this partial correction occurs at object distances with which the associated magnification varies from infinity to unity respectively. For values of $(\rho - a)/a$ greater than 0·05 the correction fails owing to the relative smallness of the quadratic term.

It may appear somewhat surprising that the optical performance of the prism deteriorates as ρ increases and therefore as the deviation of the rays decreases. That this is so can be seen by considering the angular aberration of the focusing of a parallel bundle, which can be found from (14). For very large values of ρ, e becomes $b^2/a\rho$, while the corresponding value of the deviation of a principal ray $2 . \phi_0$ becomes $2 . a/\rho$. Now the badness in the focusing performance can be judged by some simple multiple of this angular aberration relative to the deviation of the principal ray, say $1000e/2\phi_0$, and as ρ gets very large this becomes $500b^2/a^2$. Table 1 gives the absolute values of $1000e/2\phi_0$ for different values of ρ. As e is not quite symmetrical for positive and negative values of b, the average of the absolute numerical values for $b = \pm 1$ are given.

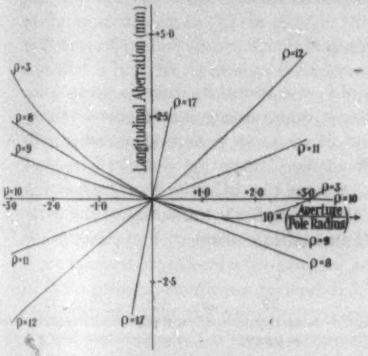


Figure 4. Homogeneous field radius 10 cm., longitudinal aberration ($f_b - f_0$)

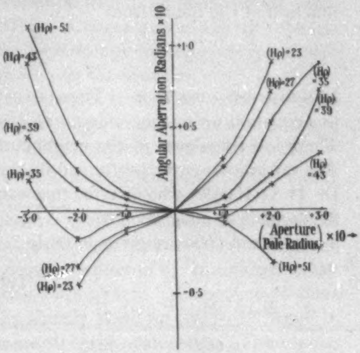


Figure 5.

Table 1

ρ (cm.)	3	8	9	10	11	12	17	50	∞
f_0 (cm.)	5·45	8·20	9·05	10	11·05	12·20	19·45	130·00	∞
Average relative angular aberration $1000\ e/2\phi_0$	0·90	0·61	0·31	0	0·32	0·67	2·00	4·56	5·00

§ 4 PRISM OF TWO POLES SEPARATED A DISTANCE EQUAL TO THEIR COMMON RADIUS

In this case θ_b can be obtained by computing in principle the path of the ray of given parameters b and $(H\rho)$ through the field of the prism. As before, a fixed pole-radius of 10 cm. was taken and the axial component of the field and its axial derivatives were obtained from tables prepared by Dr. L. J. Comrie.* The tables extend only as far as $r=32$ cm., but for values of r greater than this, H_z and its integral with respect to r can be found from the first few terms of the well known expansion in terms of Legendre polynomials.

The method of computing was to evaluate $\sin\psi$ along a trajectory using equation (8):—

$$r\sin\psi - b = \frac{1}{(H\rho)}\int_{\infty}^{r} rH_z\,dr\,.$$

The value of $\sin\psi$ given by this equation increases as r gets smaller and finally becomes greater than unity. By inverse interpolation for $\sin\psi=1$ the value of $r\psi_{\text{min}}$ (the closest real approach of the electron to the axis) is found. The value of θ at this point is the θ_b used above, and is the only value finally required. θ as a function of r is found from

$$\frac{d\theta}{dr} = \frac{1}{r}\tan\psi \qquad \ldots\ldots(17)$$

by numerical integration. Thus the value of θ_b was found by directly integrating this equation up to the value $r=r\psi_{\text{min}}$ without determining intermediate values. There is a singularity in the neighbourhood of $r\psi_{\text{min}}$, as ψ here approaches $\pi/2$. The integrations were performed by the Scientific Computing Service Ltd., and Dr. H. O. Hartley who directed this work overcame the difficulty by transforming the variable of integration to $\sqrt{r-r_{\text{min}}}$. Values of θ_b for some values of the parameters b and $(H\rho)$ are given in table 2. If this table is used with expression (9) it gives all information within its range required concerning the optical properties of the prism. The limiting focal length at infinitely small apertures f_0 is simply given by $\frac{1}{2}\,.\,db/d\theta$ at $b=0$, which can be computed by the usual methods after differencing table 2. It is quite sufficient to average the first differences, as the precise value of f_0 is of little physical importance.

Table 2

b $(H\rho)\rightarrow$	23	27	35	39	43	51
-3			0·7276	0·7697	0·8119	0·8872
-2	0·6693	0·7030	0·7767	0·8157	0·8527	0·9213
-1	0·7291	0·7601	0·8272	0·8620	0·8952	0·9569
0	0·7673	0·8160	0·8768	0·9082	0·9379	0·9934
$+1$	0·8434	0·8705	0·9258	0·9545	0·9812	1·0308
$+2$	0·8985	0·9237	0·9734	1·0003	1·0236	1·0692
$+3$		...	1·0225	1·0453	1·0667	1·1076

By examining the differences of table 2 it is found that the best focus occurs for values of $(H\rho)$ in the neighbourhood of 39, corresponding to a deviation of a principal ray through $2\phi_0=\pi-2\theta_0=1{\cdot}3252$ radians. If f_0 is taken as 10·82 cm. for this value of $(H\rho)$, the angular aberrations take the values δf given in the second row of table 3. They are clearly not symmetrical in b. One reason for such asymmetry would lie in an incorrect choice of f_0. Corresponding to a slight change of the position of focus δf a small quantity $\alpha|b|$ where $\alpha=\delta f/f_0^2$ must be added to the angular aberration. If the value of α is chosen as shown in the third row of table 3, the new angular aberration becomes within the computing error proportional to b^3, the coefficient of proportionality being slightly different according as b is positive or negative. The last two rows illustrate this proportionality.

---

* To be published.

Table 3

b	-3	-2	-1	0	$+1$	$+2$	$+3$
e	0·0039	0·0009	0·0001	0	−0·0001	0·0017	0·0067
ab	0·0006	0·0004	0·0002	0	0·0002	0·0004	0·0006
$e+ab$	0·0045	0·0013	0·0003	0	0·0001	0·0021	0·0073
$0{\cdot}00016b^3$	0·0043	0·0013	0·00016	0			
$0{\cdot}00026b^3$				0	0·00026	0·0021	0·0072

Two features of the aberrations e are of great importance. First they are of correct sign for both positive and negative values of b for correction according to (12) and they have the correct functional variation with b. The maximum value of the second term in (12) is $\frac{1}{2}(b/f_0)^3 \sim 0{\cdot}0001b^3$. Comparison of this with the last two rows of table 3 shows that it is not quite large enough for complete correction, but that it has the right sign for correction all over the aperture.

The second important feature of these aberrations is that they are of the right sign of correction with magnetic lenses, and in magnitude they are as small as, or smaller than, those of the very highest quality magnetic lenses envisaged by Rebsch (1938) in his considerations on the limit of resolution of the electron microscope.

For values of the deviation which become greater or less than the value considered above, the size of the aberration steadily increases and shows more and more marked asymmetry for positive and negative values of b. In figure 5 values of the angular aberration e derived from table 2 are plotted. Inspection of these curves shows that the asymmetry which is already small for $(H\rho)=39$ would be completely removed for the approximate value 40 of this parameter. The angular aberration would then follow a cubic law of variation with b, and the size of the coefficient C in the relation $e=C(b/f)^3$ would take the value $0{\cdot}0002f_0^{\,3} \sim 0{\cdot}25$ for both positive and negative values of b.

§ 5. PRISMS WITH DIFFERENT POLE SEPARATIONS AND FIELD SHAPES

Strictly, the properties of any field can be determined only by ray tracing in the field by some such method as that described above. It is of interest, however, to speculate how the focusing properties change as the field distribution changes by altering the pole separation, or through other causes such as general leakage from the yoke of the magnet.

For the homogeneous prism discussed above there is a complete absence of aberration in the parallel beam formed by a point object at $u=f_0=a$, over all positive and negative values of b. When the poles are separated by a distance equal to their common radius, the best image occurs for an object distance of $2f_0$, but the correcting action of the term containing uv is slightly too small. If the former prism is regarded as the limiting case of a two-pole prism with infinitely small pole separation, it may be expected that the optical properties of this class of prism would show a gradual transition as the pole separation increases. In this case, for each pole separation up to a value somewhat less than the radius of the pole, there will be one position of the object that gives a corrected conjugate image for positive and negative values of b. Each object position, corresponding to each pole separation, will only give rise to a fully corrected image for a particular value of $(H\rho)$. The deviations of the principal rays associated with these values of $(H\rho)$ will decrease slightly as the pole separation increases, the maximum deviation being $\pi/2$ for zero pole separation.

Concerning other alterations of the field distribution, little can be said except that their effect will be less than that which occurs in the corresponding problem associated with the optical properties of magnetic prisms. For the rate of deviation of a ray in the case of a prism depends directly on the field strength, whereas for the prism it depends on the axial derivative of the field.

§ 6. REAL AND VIRTUAL IMAGES

The above considerations on image formation by magnetic prisms has been based on a study of the straight rays that occur where the field strength is negligibly strong. The images are therefore in general virtual. Again, a real object placed where the field strength cannot be neglected, will have associated with it a virtual object, and it is such objects that the method used here considers. Thus this paper considers only those objects and images which are produced or observed by an external lens system, or which lie where the field strength is negligible. The aberration of a real point is not simply derivable from that of the associated virtual point unless they occur where the field is weak though not negligible. Suppose that in figure 6 P is a virtual object-point at a distance r_0 from the axis of the field. The real image associated with it will be free from aberration only if the actual trajectories of the rays belonging to it have a common solution $(r\ , \theta_1)$ for all values of the parameter b. If b is written approximately as $r_0\theta_0$, this means that $(\theta-\theta_0)$ must be constant for all values of θ_0 and one value of r. Provided that ψ is small (8) and (17) can be written

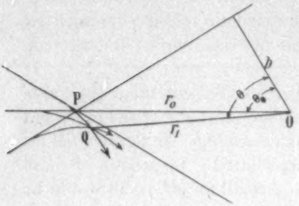


Figure 6. Real and virtual points.

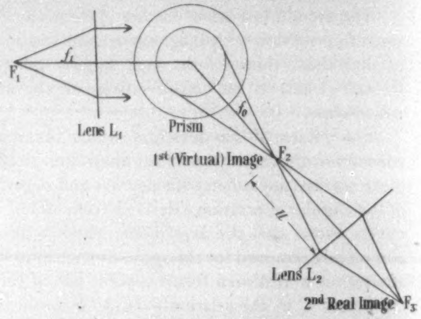


Figure 7. Prism spectrometer with two lenses.

$$-d\theta=\frac{1}{r}\left\{b+\frac{1}{(H\rho)}\int_{\infty}^{r} rH_z\,dr\right\},$$

if θ is taken increasing as r decreases according to figure 6. Thus

$$\theta = b/r + f(r),$$

where the function f is of constant form characteristic of the field distribution. Thus

$$(\theta - \theta_0) = b/r - b/r_0 + f(r).$$

This can only be constant when $r = r_0$. So that provided that $\tan\psi = \sin\psi$, the real and virtual points have the same quality but are separated by a distance $r_0 . f(r_0)$. In general these conditions are satisfied for both object and image points only when the magnification is in the neighbourhood of unity, and then only when the pole separation is not too large. The general problem of the quality of real image formation could be solved by the method of this paper provided that the integration of equation (17) was carried out step by step and the rays were fully plotted. This has not been attempted here.

§7. THE DESIGN OF PRISM β-SPECTROSCOPES

The overall design of a prism β-spectroscope can be made only when a complete knowledge of the focusing properties of prisms with respect to rays not confined to the plane of symmetry is available to supplement the study of the rays in that plane made here. Rays in the plane of the axis, for example, but not in the plane of symmetry, will be focused towards the latter, but in general this focus will be astigmatic with respect to the focus of the rays confined in that plane. Subject to these limitations a preliminary discussion of the merits of some different possible types of prism β-spectrometers can be given as follows. The prism spectrometers discussed are composed of (1) a single prism, (2) a combination of prisms, (3) a combination of a prism with axially symmetric lenses.

(1) *The single prism*

Coggershall and Muskat (1944) have considered the use of a single prism under conditions of limitingly small apertures which they effect by using a stop placed in the magnetic field in the neighbourhood of the turning points of the rays at $r_{\min}$, the rays originating from a point source. Such a design is intended for mass spectroscopy. For β-spectroscopy wide-angle focusing is desired. High resolving power will be obtained under conditions of highly corrected image formation, the conditions for which were discussed for the case of a pair of perfect poles in §§4 and 5. Since the use of a single prism requires the observation of real images, the restrictions of §6 must be satisfied. As the pair of poles are separated farther and farther the magnetic field spreads out more and more beyond the circumference of the circle of the pole $r = a$. At the same time the distance that the object has to be placed from the axis in order to obtain a corrected image increases, so that for each pole separation, a real object at the position of minimum aberration will give rise to an image of much the same quality as a virtual object would if it was placed at the same distance from the axis. In particular, corrected image formation can be achieved by working at unit magnification with a pole separation somewhat less than the pole radius. For a particular value of $(H\rho)$ corresponding to a value somewhat greater than $2.\phi_0 = 1{\cdot}3252$, for the deviation of a principal ray, the spectrometer will have a very high resolving power. The resolving power will steadily decrease as the deviation increases or decreases from this optimum value.

(2) *A combination of prisms*

Korsunsky (1945) has suggested the use of two prisms in series with the object at the focal point of the first, and the image at the focal point of the second. This appears to have no advantage over the use of a single prism. Under the conditions proposed by him for its use, the angular aberrations in the first prism are added to those produced by the second. Thus the increase of dispersion over that due to a single prism is not accompanied by a corresponding increase of resolving power. The aberrations produced in one prism could be made to correct those produced in the other if the direction of deviation was chosen for the second prism so that the parameters b for any ray passing through the combination took different signs in the two prisms. But under these conditions the image formation becomes nearly achromatic, so that the dispersion and resolving power disappear. Korsunsky's four-pole spectrometer was developed from a consideration of rays in the plane of symmetry taking the case of zero pole separation. On these grounds advantage from the use of two prisms would only accrue if the quality of image formation for small pole separations and high magnification was superior to that at unit magnification and much larger pole separation. The total focusing of a cone of rays by each prism will be astigmatic and the possibility does remain of using a system of prisms to produce a pair of stigmatic points.

(3) *Combination of a prism with lenses*

Klemperer (1935) described a spectrometer that he constructed and that was modelled roughly on the lines of an optical prism spectrometer. Unfortunately this work had to be abandoned before the best working conditions were found. Now that the focusing action of the prism itself are more clearly realized, an instrument of this type could be designed as shown in figure 7. Here the magnetic lens L_1 forms a parallel bundle which is deviated by the prism and brought to a focus at F_2. In general this image will be virtual and must be observed with another lens L_2. Under certain conditions the aberrations of the prism have opposite sign to those of the lenses, so that corrected image formation at F_3 would be possible over a wide aperture for one value of $(H\rho)$ provided that lenses of sufficiently high quality were available. Unfortunately there is considerable uncertainty in this matter. It is known that the angular aberration introduced by a magnetic lens is proportional to $(b/f)^3$, where b is the intercept of the ray in the plane of the lens, regarded as thin, although it is not clear that this functional variation will hold all over the large apertures considered here. If C_1 is the coefficient of proportionality, so that for a lens the angular aberration becomes $C\,(b/f)^3$, and if C_2 is the corresponding coefficient for the prism, the total angular aberration of the final image of figure 7 is $C\,(b/f_1)^3 - C_2(b/f_0)^3 + C_1(b'/f_2)^3$ where $b' = b(u/f_1)^3$ and f_1 is the focal length of the first lens, f_0 that of the prism and f_2 that of the second lens. Thus for corrected image formation:—

$$C_1/f_1{}^3 - C_2/f_0{}^3 + C_1/(f_0 f_2/u)^3 = 0.$$

For a real final image, u must be greater than f_2 so that for this equation to hold C_2 must be greater than C_1. From the discussion given in §4 it would appear that values of the order of 0·25 can be attained for C_2. For the type of lenses used in β-ray spectroscopy the values of C_2 are not available. From simple theoretical considerations the writer (1942) showed that for a class of short solenoids advantage was to be expected by a reduction of the ratio (diameter/length) which reduction fortunately also favours economy of power. However, the C_1 values have not yet been worked out. In connexion with the electron microscope, Rebsch suggested a lower limit of $C_1=0{\cdot}25$, but his argument was based on considerations of instrumentation that do not necessarily apply to the lenses required for β-spectroscopy. More recently slightly smaller values have been suggested.

Compared with the single prism, the compound spectrometer employing lenses suffers from two disadvantages, in that it is more complicated and that the correction is a much more formidable task. The final choice between the two types depends ultimately on the C values attainable in magnetic lenses, and the astigmatic focusing properties of the prism.

ACKNOWLEDGMENTS

I should like to thank Prof. J. D. Bernal, F.R.S., for his encouragement and generous support. My thanks are due also to Dr. L. Jánossy for discussion concerning the method of ray tracing, to Drs. Comrie and Hartley, of the Scientific Computing Service Ltd., for discussion concerning the numerical integrations, and to Dr. Ehrenberg for general discussions.

REFERENCES

Coggershall and Muskat, 1944. *Phys. Rev.*, **66**, 189.
Herzog, 1934. *Z. Phys.*, **89**, 447.
Klemperer, 1935. *Phil. Mag.*, **20**, 545.
Korsunsky, 1945. *J. Phys. U.S.S.R.*, **9**, 14.
Rebsch, 1938. *Ann. Phys., Lpz.*, **31**, 551.
Sawyer, 1936. *Proc. Camb. Phil. Soc.*, **32**, 453.
Siday, 1942. *Proc. Phys. Soc.*, **54**, 226.

THE THEORY OF RADIATION DAMPING

By J. HAMILTON,
University of Manchester

MS. received 3 March 1947

§1. INTRODUCTION AND SUMMARY

This paper is the result of an attempt to get a better understanding of the quantum theory of radiation damping developed by Weisskopf and Wigner (1930), Heitler (1941), Wilson (1941) and Heitler and Peng (1942). It has been suggested by Peng (1944) that the Heitler-Peng damping equation arises because of the degeneracy of the unperturbed states of the system of particles and radiation; and he has claimed (1946) that a careful mathematical treatment of the "quasi-degenerate" continuum of energy levels not only gives the damping equation, but also removes the well known divergence difficuties of radiation theory.

It is obvious that the concept of degeneracy is much simpler when the system is enclosed in a box, and as a consequence, the energy levels are discrete. The difficulties of the solution then lie rather in the complexity of the mathematics than in the concepts. It has, however, been possible, subject to certain limitations, to solve the emission and scattering problems exactly for a cubic box in the discrete energy case. These limitations are just those imposed in the original Dirac and Heitler-Peng methods of calculating the transition probabilities—namely the omission of states which seem to be physically unimportant. That this limitation is a very severe one, can be seen from the one case in which it has been possible to consider more than the conventional states. The addition of some physically unimportant states to the equations of the emission problem reduces the divergence difficulties. In general, however, it seems impossible to use the discrete energy method to investigate whether or not the divergencies in radiation theory are purely mathematical difficulties due to the neglect of higher-order processes. It is as well, therefore, to make it clear at the outset that it is really models of the actual systems which are considered. For example, the scattering problem corresponds to a large number of oscillators, whose energies are close together, which are coupled to each other through certain intermediate oscillators whose energies are very different from those of the main set.

The questions which then arise are (i) what rôle degeneracy plays in the solution, (ii) why the perturbation method does not give the correct transition probabilities in scattering problems, (iii) how the behaviour of such a model differs from the actual radiation system as observed. Degeneracy, in the strict sense, plays no rôle whatever, as the enclosing box can be chosen so that the ratios of the squares of the edges are irrational numbers. In that case there is no degeneracy. The perturbation method, however, breaks down for strong interactions in scattering problems. A rough criterion can be given when the compound interaction matrix elements * are not strongly dependent on the directions of the incident and scattered particles. Then the product of the average magnitude of the matrix elements with the density function † $\rho(E)$ of the scattered states must be much less than unity ($\hbar=1$) for the perturbation procedure to be applicable. What is happening can easily be seen when the energy levels of the unperturbed system are highly degenerate. The interaction causes a splitting of each energy level, and it is only when the splitting corresponds to a change in energy which is small compared with the energy difference between two adjacent levels that the perturbation method is applicable. For a sufficiently strong interaction, the first-order energy perturbation can be made very large; but the exact solution shows that the levels which split off from a degenerate level E_2 cannot go further than roughly half-way between this degenerate level and the adjacent degenerate levels E_1 and E_3. An exact criterion for the validity of the perturbation method is that the magnitudes of all the eigenvalues of a homogeneous integral equation must be much greater than unity. The kernel of this homogeneous integral equation is the kernel of the Heitler-Peng integral equation; and the two equations are closely related.

The chief difference between the behaviour of the model and the observed

---

* In the normal sense.

† Which also has to be little dependent on the angles.

behaviour of the atomic emitter is the line shift. The final states of the model assume the usual energy distribution or line form; but the centre of the line does not coincide with the initial energy. The shift is very large or infinite. However, the energy *is* conserved. The reason for this strange situation can be seen as follows. The usual line form centred about the resonant frequency gives a small probability for the emission of very high-energy quanta. When the total emitted energy is calculated, it is seen to diverge due to these high-energy quanta. The shift in the line compensates for this effect. The shift is very much reduced if a set of states, which are of no physical importance, is introduced. The usual states are (i) the atom in its excited state, (ii) the atom in its normal state and one photon present; and in the new states (iii) the atom is in its excited state and two photons are present. The very improbable transitions from (ii) to (iii) very largely counteract the shift. In the scattering case the comparison with reality is more complicated. Several terms of the first-order self-energy type $\sum_r |H_r|^2/(E_0-E_r)$ involving both the simple and the compound matrix elements, have to be neglected. When this is done the transition probabilities are identical with the Heitler-Peng result.

It is assumed both in the emission and the scattering cases that terms of the type $\sum_r |H_r|^2/(E_0-E_r)^2$ lead to no difficulties. For this to be true it is only necessary that the contribution to the sum from the region of large (E_r-E_0) is finite. Even when the number of oscillators tends to infinity the sum remains finite, provided a cut-off is made at any energy, however high. (For example, an electron whose radius is 10^{-100} cm. would suffice.)

§2. THE EIGENVALUE PROBLEM

The motion of a number of coupled oscillators can be treated by the transition probability or by the eigenvalue method. The latter method is used here.

Suppose that on neglecting the coupling, the N component parts of the system are described by a Hamiltonian H°, which leads to the N stationary states described by the wave functions $\psi_1, \psi_2, \ldots \psi_N$; whose energies are $E_1, E_2, \ldots E_N$ respectively. These wave functions can be written as

$$\psi_r = \phi_r e^{-iE_r t} \qquad (r=1, 2, \ldots N),$$

where the function ϕ_r is independent of time (and $\hbar=1$). Any state of the system is described by

$$\Psi = \sum_{r=1}^{N} a_r \psi_r = \sum_{r=1}^{N} a_r \phi_r e^{-iE_r t}, \qquad \ldots\ldots(1)$$

the a_r being constants when there is no coupling.

Introducing a time-independent interaction V between the components of the system, the equation of motion becomes

$$i\frac{\partial \Psi}{\partial t} = (H^\circ + V)\Psi. \qquad \ldots\ldots(2)$$

If the ϕ_r are chosen as a normal orthogonal set such that *

$$(\overline{\phi_r} \cdot \phi_s)_v = \delta_{rs},$$

---

* $(\ldots\ldots)_v$ represents the integral over the volume in which the system is contained.

then (2) gives

$$i\frac{d}{dt}(a_r e^{-iE_r t}) = \sum_{s=1}^{N} H_{rs} a_s e^{-iE_s t}, \qquad \ldots\ldots(3)$$

where $H_{rs} = (\bar{\phi}_r \,.\, H_0 + V \,.\, \phi_s)_v$, and is independent of the time.

The set (3) can be solved in terms of a series of N "normal modes" of amplitudes A_r and energies Λ by substituting

$$a_r e^{-iE_r t} = A_r e^{-i\Lambda t}.$$

This gives

$$A_r \,.\, \Lambda = \sum_{s=1}^{N} H_{rs} A_s, \qquad \ldots\ldots(4)$$

and hence the Λ are the roots of

$$|H_{rs} - \Lambda \,.\, I| = 0. \qquad \ldots\ldots(5)$$

H_{rs} being Hermitian, the N roots Λ_μ ($\mu = 1, 2, \ldots\ldots N$) are real; and denoting the amplitudes of the μth mode by A_r^μ, the normal modes are

$$\Psi^\mu = \sum_{r=1}^{N} a_r^\mu \psi_r, \quad \text{where} \quad a_r^\mu = A_r^\mu \,.\, e^{-i(\Lambda_\mu - E_r)t}.$$

The A_r^μ can satisfy the conditions

$$\sum_{r=1}^{N} \bar{A}_r^\nu A_r^\mu = \delta_{\mu\nu} \qquad (\mu, \nu = 1, 2, \ldots\ldots N),$$

and it follows that

$$\sum_{\nu=1}^{N} \bar{A}_r^\nu A_s^\nu = \delta_{rs} \qquad (r, s = 1, 2, \ldots\ldots N). \qquad \ldots\ldots(6)$$

The general state of the coupled system can be written

$$\Psi = \sum_{\mu=1}^{N} c_\mu \Psi^\mu, \quad \text{where} \quad \sum_{\mu=1}^{N} |c_\mu|^2 = 1;$$

so from (1)

$$a_r e^{-iE_r t} = \sum_{\nu=1}^{N} c_\nu A_r^\nu e^{-i\Lambda_\nu t}.$$

If the initial conditions are $a_r(0)$ ($r = 1, 2, \ldots\ldots N$), then

$$a_r(0) = \sum_{\nu=1}^{N} c_\nu A_r^\nu$$

or

$$c_\nu = \sum_{r=1}^{N} \bar{A}_r^\nu a_r(0) \qquad (\nu = 1, 2, \ldots\ldots N). \qquad \ldots\ldots(7)$$

The A_r^ν are merely the coefficients of the unitary matrix **S** which transforms the system from the representation in which H° is diagonal to that in which $H^\circ + V$ is diagonal, viz.: $S_{ik} = A_i^k$.

The conservation and spread of the energy of the oscillators due to the interaction can be readily deduced. The average energy $\bar{E}$ of a state $\bar{\Psi}$ is

$$\bar{E} = \left(\bar{\Psi},\, i\frac{\partial}{\partial t}\Psi\right)_v.$$

This gives

$$\overline{E}=\sum_{\mu=1}^{N}|c_\mu|^2\Lambda_\mu,$$

and similarly

$$\overline{E^2}=\sum_{\mu=1}^{N}|c_\mu|^2\Lambda_\mu^2.$$

If the initial conditions are

$$\left.\begin{aligned}a_k(0)&=1,\\ a_r(0)&=0,\quad r\neq k,\end{aligned}\right\}\qquad\ldots\ldots(8)$$

then $c_r=\overline{A}{}^r_k$, so

$$\overline{E}=\sum_{\mu=1}^{N}|A^\mu_k|^2\,.\,\Lambda_\mu=\sum_{\mu=1}^{N}\overline{A^\mu_k}\sum_{s=1}^{N}H_{ks}\,A^\mu_s=\sum_{s=1}^{N}H_{ks}\delta_{ks}.$$

Thus

$$\overline{E}=E_k+V_{kk}.\qquad\ldots\ldots(9)$$

Similarly

$$\overline{E^2}=\sum_{r=1}^{N}|H_{kr}|^2,$$

and

$$\overline{(E-E_k)^2}=\sum_{r=1}^{N}|V_{kr}|^2.\qquad\ldots\ldots(10)$$

§ 3. THE FUNDAMENTAL EQUATIONS

(a) *Atomic emission*

The system consists of an atom enclosed in a box with perfectly reflecting walls. Following the usual treatment, the states considered are:—

ψ_0—the atom in its excited state of energy E_0 with no photons present.

ψ_r—the atom in its normal state of energy E, and a photon of energy ϵ_r present.

It is assumed that there are N states of the latter kind. The interaction V allows single quantum jumps only, i.e. $V_{rs}=0$ unless one of r or s is zero. Putting $V_{0r}=H_{0r}$, the fundamental equations (4) connecting the $N+1$ states ψ_i $(i=0, 1, \ldots N)$ become

$$\left.\begin{aligned}&(E_0-\Lambda)A_0+\sum_{s=1}^{N}H_{0s}A_s=0,\\ &(E+\epsilon_r-\Lambda)A_r+H_{r0}A_0=0\qquad(r=1,\,2,\,\ldots\,N).\end{aligned}\right\}\ldots\ldots(11)$$

If two or more energies ϵ_r are equal, degenerate solutions will exist. Let $\epsilon_{r_1}=\epsilon_{r_2}=\ldots=\epsilon_{r_m}$ be degenerate levels. Then the degenerate solution satisfies

$$\left.\begin{aligned}&H_{r_i0}A_0=0\qquad(i=1,2,\ldots m),\text{ i.e. }A_0=0,\\ &A_s=0\qquad\epsilon_s\neq\epsilon_{r_i},\\ &\sum_{i=1}^{mr}H_{0r_i}A_{r_i}=0.\end{aligned}\right\}\qquad\ldots\ldots(12)$$

The last conditions shows that the vector $(A_{r_1},\ldots A_{r_m})$ must be orthogonal to the vector $(\overline{H}_{0r_1},\ldots\overline{H}_{0r_m})$, so there are $(m-1)$ independent degenerate solutions.

If r denotes the various degenerate levels, the level ϵ_r being m_r-fold and Λ_μ ($\mu=0$, $1, \ldots M$) the remaining roots such that $\Lambda_\mu \neq E+\epsilon_s$ for any s, it might be thought that the general solution of equation (3) should be

$$a_k e^{-iE_k t} = \sum_{\mu=0}^{M} A_k^\mu e^{-i\Lambda_\mu t} + \sum_{r} \sum_{\beta=1}^{m_r-1} A_k^\beta \left(\frac{t}{i}\right)^{\beta-1} e^{-i(E+\epsilon_r)t}.$$

On substituting this expression in (3) and using relations (12), it can be seen that the A_k^β vanish except when $\beta=1$, so no powers of t occur in the solution. For the emission problem the initial conditions are

$$\left.\begin{aligned} a_0(0)&=1, \\ a_r(0)&=0 \qquad (r=1, 2, \ldots N). \end{aligned}\right\} \qquad \ldots\ldots(13)$$

Then $c_\mu = \bar{A}_0^\mu$; and if $\Lambda_\nu = E+\epsilon_r$ then $c_\nu=0$. So such degenerate states do not enter into the solution.

To find the non-degenerate solutions, multiply the first equation (11) by $\prod_{r=1}^{N} (E+\epsilon_r-\Lambda)$, and substitute for A_s, using the second equation (11), giving

$$\left\{(E_0-\Lambda) \prod_{r=1}^{N} (E+\epsilon_r-\Lambda) - \sum_{s=1}^{N} \prod_{r=1}^{N}{}_{(s)} (E+\epsilon_r-\Lambda)\, |H_{0s}|^2\right\} A_0 = 0\,^*$$

Thus if $A_0 \neq 0$, and if there are m_r states with energy $E+\epsilon_r$, division by

$$\prod_{r=1}^{N} (E+\epsilon_r-\Lambda)^{m_r-1}$$

gives

$$\Lambda - E_0 + \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(E+\epsilon_r-\Lambda) \right\} = 0, \qquad \ldots\ldots(14)$$

where M is the number of distinct ϵ_r values and r_i ($i=1, 2, \ldots m_r$) refer to the states with energy $E+\epsilon_r$.

The function

$$f(\Lambda) = \Lambda - E_0 + \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(E+\epsilon_r-\Lambda) \right\}$$

has simple poles at $\Lambda = E+\epsilon_r$, and its derivative

$$\frac{df(\Lambda)}{d\Lambda} = 1 + \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(E+\epsilon_r-\Lambda)^2 \right\}$$

has poles of order 2 at $\Lambda = E+\epsilon_r$, and for real $\Lambda \dfrac{df(\Lambda)}{d\Lambda} > 1$. Hence the equation $f(\Lambda)=0$ has one solution between $E+\epsilon_r$ and $E+\epsilon_{r+1}$ ($r=1, 2, \ldots M-1$) and the 1st and ($M+1$)th solutions are less than $E+\epsilon_0$, and greater than $E+\epsilon_M$, respectively, these being the least and the greatest energy levels†. Figure 1 shows the graphical solution.

If $\sum_{s=1}^{N} |H_{0s}|^2/(E+\epsilon_s) \to \infty$ as $N \to \infty$, the lowest root $\Lambda \to -\infty$, and the root between $E+\epsilon_r$ and $E+\epsilon_{r+1}$ tends to $E+\epsilon_r$ as the limit. Thus N must in general be kept finite until the solution is obtained.

---

* $\prod_{r=1}^{N}{}_{(s)} = \prod_{r=1}^{s-1} \cdot \prod_{r=s+1}^{N}$.

† A footnote to Weisskopf and Wigner's (1930) paper shows that they were aware of equation (14) and its chief properties.

(b) *Scattering*

The scattering of a photon or a meson by electrons or nucleons is considered. The system consists of the scattering particle enclosed in a box so that a large number of different photon or meson states can occur. Initially there is one photon or meson, and only those scattered states in which one quantum is present are allowed. Intermediate states can occur, but it is necessary to impose the condition that the energy of any intermediate state is far removed from the energies of the initial and the final states—a physically reasonable assumption. The interaction can lead to either one or two quantum jumps.

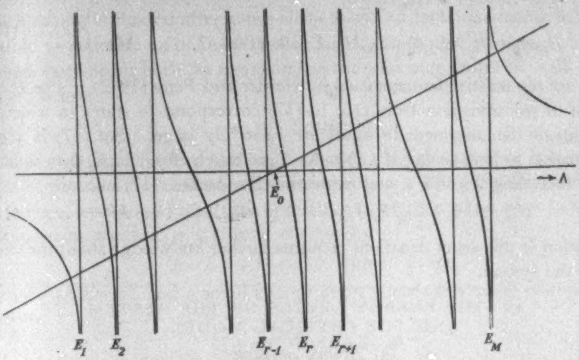


Figure 1. The intersections of the functions $\Lambda - E_0$ and $\sum_{r=1}^{M} |H_r|^2/(\Lambda - E_r)$.

Denoting the initial, intermediate and final states by the subscript indices 0; i, j; r, s; the fundamental equations (4) become

$$\left.\begin{aligned}
&(E_0-\Lambda)A_0+\sum_i H_{0i}A_i+\sum_s H_{0s}A_s=0,\\
&H_{i0}A_0+(E_i-\Lambda)A_i+\sum_s H_{is}A_s=0,\\
&H_{s0}A_0+\sum_i H_{si}A_i+(E_s-\Lambda)A_s+\sum_r{}_{(s)} H_{sr}A_r=0^*.
\end{aligned}\right\}\ldots\ldots(15)$$

If there are P intermediate and M final states, this set of equations gives $(1+P+M)$ roots Λ on eliminating all the A. It is more convenient to eliminate the A_i first, giving:—

$$\left.\begin{aligned}
&\left\{E_0-\Lambda-\sum_i \frac{H_{0i}H_{i0}}{E_i-\Lambda}\right\}A_0+\sum_s\left\{H_{0s}-\sum_i \frac{H_{0i}H_{is}}{E_i-\Lambda}\right\}A_s=0,\\
&\left\{H_{s0}-\sum_i \frac{H_{si}H_{i0}}{E_i-\Lambda}\right\}A_0+\left\{E_s-\Lambda-\sum_i \frac{H_{si}H_{is}}{E_i-\Lambda}\right\}A_s\\
&\qquad+\sum_r{}_{(s)}\left\{H_{sr}-\sum_i \frac{H_{si}H_{ir}}{E_i-\Lambda}\right\}A_r=0\quad(s=1,2,\ldots M).
\end{aligned}\right\}\ldots\ldots(16)$$

---

* $\sum_r{}_{(s)}$ denotes the summation over all final states except s.

The terms $\sum_i \frac{|H_{0i}|^2}{E_i-\Lambda}$, $\sum_i \frac{|H_{si}|^2}{E_i-\Lambda}$ are clearly second-order self-energy terms. Further, the remaining summations containing $E_i-\Lambda$ in the denominator in general only contain a finite number of terms. If only those values of Λ which are in the vicinity of the initial energy are to be considered, Λ can be replaced in these summations by E_0 (as then $|E_i-\Lambda| \geqslant |E_0-\Lambda|$). Writing

$$E'_s = E_s - \sum_i \frac{|H_{si}|^2}{E_i-\Lambda}, \qquad H'_{sr} = H_{sr} - \sum \frac{H_{si}H_{ir}}{E_i-E_0}, \text{ etc.}$$

the set (16) becomes

$$\left.\begin{aligned} &(E'_0-\Lambda)\, A_0 + \sum_s H'_{0s}\, A_s = 0, \\ &H'_{s0}A_0 + (E'_s-\Lambda)A_s + \sum_r{}_{(s)} H'_{sr}A_r = 0. \quad (s=1, 2, \ldots M). \end{aligned}\right\} \quad \ldots\ldots(17)$$

The H'_{sr} are the matrix elements used by Heitler and Peng (1942). The P roots neglected in the transition from (16) to (17) correspond to states in which the amplitudes of the intermediate states are relatively large. Set (17) is a good approximation so long as its $(M+1)$ roots Λ are much closer to E_0 than to any of the E_i. Including 0 in the s, and neglecting the dashes, (17) becomes

$$(E_s-\Lambda)A_s + \sum_r{}_{(s)} H_{sr}A_r = 0 \qquad (s=0, 1, 2, \ldots M) \qquad \ldots\ldots(18)$$

The solution of this set of equations requires further knowledge about the energy levels of the system.

§4. THE ENERGY LEVELS OF THE SYSTEM AND THE PHYSICAL MODEL

(a) *Atomic emission*

The box enclosing the system can be taken as a cube with sides of length L. A radiation oscillator of vector momentum $\boldsymbol{p}$ is described by a wave function

$$\phi_p = L^{-3/2} e^{-i(\mathbf{p}.\mathbf{x})/\hbar}$$

where $\boldsymbol{x}$ is the position vector. The following periodicity conditions are imposed:

$$p_x \,.\, L = \boldsymbol{h} \,.\, n_x,$$
$$p_y \,.\, L = \boldsymbol{h} \,.\, n_y,$$
$$p_z \,.\, L = \boldsymbol{h} \,.\, n_z,$$

where n_x, n_y, n_z are integers, positive, negative or zero. Thus

$$p = \frac{\boldsymbol{h}}{L}\sqrt{n_x^2 + n_y^2 + n_z^2} = \frac{\boldsymbol{h}}{L}\sqrt{n}, \text{ say.}$$

For some values of the positive integer n, the equation

$$n_x^2 + n^2 + n_z^2 = n \qquad \ldots\ldots(19)$$

has no solutions, but in general it has many. The average number of sets (n_x, n_y, n_z) which give a value of $\sqrt{n_x^2 + n^2 + n_z^2}$ lying in the range $(\sqrt{n}, \sqrt{n}+\delta)$ is $4\pi n \,.\, \delta$. For large n,

$$\sqrt{n+1} - \sqrt{n} \simeq 1/2\sqrt{n},$$

so the average number of solutions of (19) for each n is $2\pi\sqrt{n}$. The average number of states with the same momentum p is $2\pi\sqrt{n} = 2\pi \,.\, pL/\boldsymbol{h}$, and the difference between adjacent values of p is, approximately, $(\boldsymbol{h}/L) \,.\, 1/2\sqrt{n} = (\boldsymbol{h}/L)^2/2p$, or a small integral multiple thereof.

For photons, the energy ϵ is given by $\epsilon = \mathbf{c}p$, so the average number of states of the same energy ϵ is $2\pi\epsilon \,.\, L/\mathbf{c}h$ and the separation of adjacent energy levels is $(\mathbf{c}h/L)^2/2\epsilon$, or a small integral multiple of it. [The total number of states in the energy range $(\epsilon, \epsilon + d\epsilon)$ is the well known expression $4\pi\epsilon^2 d\epsilon \,.\, (L/\mathbf{c}h)^3$.]

It is clear that the energy levels are mostly highly degenerate, and that the distinct energy levels lie very close together. If the total number of oscillators N is kept finite, the atomic emission problem, developed in § 3 above, can be solved approximately, as it stands. Comparing the series

$$\sum_{\mu=0}^{M} \frac{|A^\mu|^2}{(E+\epsilon_r-\Lambda)^2} \quad \text{and} \quad \sum_{\mu=0}^{M} \frac{1}{(E+\epsilon_r-\Lambda)^2}$$

(which can be evaluated after some elementary but laborious manipulation) it is possible to estimate $|A^\mu_r|^2$, and hence find the complete solution. A simpler method, suitable for both the emission and the scattering problems, will, however, be used below.

(b) *Scattering*

If the momenta of the incident radiation and the scatterer are equal and opposite, the total momentum will always be zero. For simplicity, consider the scattering of a photon by a particle of rest mass $\mu/\mathbf{c}^2$. In a final state in which the photon has momentum p, the total energy is

$$E = k + \sqrt{k^2+\mu^2}, \quad \text{where} \quad k = p\mathbf{c}.$$

Thus $E - \mu^2/E = 2k$, and a small change δk in k leads to a small change δE in E given by

$$(1 + \mu^2/E^2) \,.\, \delta E = 2 \,.\, \delta k.^* \qquad \ldots\ldots(20)$$

So the distance between adjacent energy levels is

$$\Delta E = (\mathbf{c}h/L)^2 \,.\, \frac{1}{k(1+\mu^2/E^2)}$$

or a small integral multiple thereof; and the average degree of degeneracy of each level is†

$$\begin{aligned} m &= 2\pi k \,.\, (L/\mathbf{c}h), \\ &= \pi(E - \mu^2/E) \,.\, (L/\mathbf{c}h). \end{aligned}$$

If the intervals between adjacent energy levels were equal in the vicinity of the initial energy value and if all energy levels in this region had the same degree of degeneracy, the solution for the emission and the scattering cases would be very much simplified. In the emission case the line breadth has a finite value irrespective of the size of the enclosing box, so it is the average behaviour of the A^μ_r over a large number of energy levels which is important. Therefore it is not be expected that the change in the solution due to altering the energy levels to make them equally degenerate will be of any practical importance. The scattering problem will also be treated for such a physical model in which the energy levels are equally spaced, and all are equally degenerate (in the vicinity of the initial

---

* For the scattering of a meson of rest mass $\mu'/\mathbf{c}^2$ by a particle of rest mass $\mu/\mathbf{c}^2$ equation (20) becomes

$$\{E^4 - (\mu^2 - \mu'^2)^2\} \,.\, \delta E = 4kE^3 \,.\, \delta k.$$

† Of course there may be a further degeneracy for given energy and direction due to spin, polarization, etc.

energy). The transition probabilities can only depend on some function of the interaction matrix $\boldsymbol{H}$ and on the density of the final states over a small region which may contain as many energy levels as is desired.

§ 5. THE SOLUTION OF THE EMISSION PROBLEM

It is further assumed that, over the small energy range about the initial energy value E_0, in which the energy levels are taken to be equally spaced and equally degenerate, the matrix elements H are independent of the energy of the photons. The error thus introduced will be seen to be negligible. Besides the equation

$$\Lambda - E_0 + \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(E+\epsilon_r-\Lambda) \right\} = 0; \qquad \ldots\ldots(14)$$

the normalizing condition for any solution, viz.:—

$$|A^\mu|^2 \left[1 + \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(E+\epsilon_r-\Lambda)^2 \right\} \right] = 1. \qquad \ldots\ldots(21)$$

is required. This is merely the equation $|A_0^\mu|^2 + \sum_{r=1}^{N} |A_r^\mu|^2 = 1$. Put $E_r = E + \epsilon_r$. The position of the root Λ which lies in the vicinity of some energy value E' is obtained as follows:—

$$\begin{aligned} \Lambda - E_0 &= \sum_{r=1}^{M} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(\Lambda - E_r) \right\} \\ &= \sum_{r=s-n}^{s+n} \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(\Lambda - E_r) \right\} + \left(\sum_{r=1}^{s-n-1} + \sum_{r=s+n+1}^{M} \right) \left\{ \sum_{i=1}^{m_r} |H_{0r_i}|^2/(\Lambda - E_r) \right\}, \end{aligned}$$

where $E' = E_s$, and n is a large positive integer. Further, the summation over the degenerate states of any energy level is equivalent to an angular integration, thus

$$\sum_{i=1}^{m_r} |H_{0r_i}|^2 = \frac{m_r}{4\pi} \int |H_{0r_i}|^2 \, d\Omega_i. \qquad \ldots\ldots(22)$$

The first term in the expression above for $\Lambda - E_0$ can be written

$$\sum_{i=1}^{m} |H_{0s_i}|^2 . \sum_{r=-n}^{+n} \frac{1}{(\Lambda - E_s) - r\Delta E},$$

where ΔE is the interval between adjacent energy levels in the vicinity of E_s.

Further,

$$\sum_{r=-\infty}^{+\infty} \frac{1}{(\Lambda - E_s) - r\Delta E} = \lim_{n\to\infty} \sum_{r=-n}^{+n} \frac{1}{(\Lambda - E_s) - r\Delta E} = (\pi/\Delta E) \, . \, \cot\{(\Lambda - E_s)\pi/\Delta E\}$$

so

$$\begin{aligned} \Lambda - E_0 = &\sum_{i=1}^{m} |H_{0s_i}|^2 . (\pi/\Delta E) . \cot\{(\Lambda - E_s)\pi/\Delta E\} \\ &+ \int_{(E_s)} (E_s - E_r)^{-1} . \int |H_{0r}|^2 \, d\Omega \, . \, \rho(E_r) dE_r + \eta, \qquad \ldots\ldots(23) \end{aligned}$$

where η is a small error and (E_s) denotes that the range $(E_s - \xi, E_s + \xi)$, where $\xi = n\Delta E$, is excluded from the integration. $\rho(E) . dE . d\Omega$ is the total number of states of energy lying in the range $(E, E + dE)$, and with momenta lying in the angle $d\Omega$. ξ can be given any finite value, e.g. $0{\cdot}00001 \times E'$, and then if the size of the box is sufficiently large the error η can be made as small as desired, compared with the first term on the right-hand side of (23).

Putting

$$R(E_s)=\int_{(E_s)}(E_s-E_r)^{-1}\,.\int|H_{0r}|^2\,d\Omega\,.\,\rho(E_r)\,dE_r$$

the equation for Λ becomes

$$\Lambda-E_0=\pi\,.\int|H_{0s}|^2\,d\Omega\,.\,\rho(E_s)\,.\,\cot\{(\Lambda-E_s)\pi/\Delta E\}+R(E_s)\quad\ldots\ldots(24)$$

Equation (21) is treated similarly, viz.:—

$$\sum_{r=1}^{M}\left\{\sum_{i=1}^{mr}|H_{0r_i}|^2/(\Lambda-E_r)^2\right\}=\sum_{i=1}^{m}|H_{0s.}|^2\,.\sum_{r=-n}^{+n}\frac{1}{(\Lambda-E_s-r\Delta E)^2}$$
$$+\int_{(E_s)}(E_s-E)^{-2}\,.\int|H_{0r}|^2\,d\Omega\,.\,\rho(E)\,dE,$$

and

$$\sum_{r=-\infty}^{+\infty}\frac{1}{(\Lambda-E_s-r\Delta E)^2}=(\pi/\Delta E)^2\,.\,\mathrm{cosec}^2\{(\Lambda-E_s)\pi/\Delta E\};$$

so

$$\sum_{r=1}^{M}\left\{\sum_{i=1}^{mr}|H_{0r_i}|^2/(\Lambda-E_r)^2\right\}=\pi\rho(E_s)\int|H_{0s}|^2\,d\Omega\,.\,(\pi/\Delta E)\,\mathrm{cosec}^2\{(\Lambda-E_s)\pi/\Delta E\}$$
$$+\int_{(E_s)}(E_s-E)^{-2}\int|H_{0r}|^2\,d\Omega\,.\,\rho(E)\,dE+\eta'.\qquad\ldots\ldots(25)$$

The first term on the right-hand side of (25) increases like $1/\Delta E$ as L increases, and η' the error becomes small compared with the first term for sufficiently large L. The remaining term is independent of L, and, provided the integral is finite, it becomes negligible compared with the first term for sufficiently large L. It is of interest to note that the integral may have any finite value however large. Even for unbound states this could be satisfied by taking a cut-off value (electron radius) of 10^{-100} cm., say.

Thus we have (using (24)

$$|A_0^\mu|^2\,.\left[1+\pi\rho(E_s)\,.\int|H_{0s}|^2\,d\Omega\,.\,(\pi/\Delta E)\,\mathrm{cosec}^2\{(\Lambda-E_s)\pi/\Delta E\}\right]=1$$

and

$$|A_0^\mu|^2\,.\,\frac{1}{\Delta E}\to\frac{\rho(E_s)\int|H_{0s}|^2\,d\Omega}{\left\{\pi\rho(E_s)\int|H_{0s}|^2d\Omega\right\}^2+\{E_s-E_0-R(E_s)\}^2}\qquad\ldots\ldots(26)$$

as $L\to\infty$. $E_s-E_0-R(E_s)$ is substituted for $\Lambda_\mu-E_0-R(E_s)$ as the root Λ_μ is close to E_s.

Further,

$$a_0e^{-iE_0t}=\sum_{\mu=0}^{M}|A_0^\mu|^2\,.\,e^{-i\Lambda_\mu t}$$

when the initial conditions are

$$\left.\begin{aligned}&a_0(0)=1,\\&a_r(0)=0\qquad(r=1,\,2,\,\ldots\,N).\end{aligned}\right\}\ldots\ldots(13)$$

so

$$a_0e^{-iE_0t}\simeq\int_{E_1}^{E_M}|A_0^\mu|^2\,.\,\frac{1}{\Delta E}\,.\,e^{-iE_st}\,dE_s,\qquad\ldots\ldots(27)$$

and therefore (26) gives the form of the emitted line. This expression agrees with that of Weisskopf and Wigner when the term $R(E_s)$ is neglected. If this term were independent of E_s, equation (26) would give the usual form of line centred around E_0+R. However, for values of E close to the lower limit of the energy range, $R(E)$ is negative, while it is positive close to the upper limit of the energy range. For practical cases it is large and negative in the region of E_0. A rough estimate got by assuming

$$\begin{aligned}\int |H_{0r_i}|^2\, d\Omega &= \text{const.} \quad \epsilon_r<\epsilon' \\ &=0 \qquad \epsilon_r>\epsilon',\end{aligned}$$

where $\epsilon' = \boldsymbol{hc}/a_0$ (a_0 = Bohr radius), gives a shift of the centre of the line of the order of $10^6 . \Gamma\boldsymbol{h}$ in the direction of decreasing energy, where $\Gamma\boldsymbol{h}$ is the line breadth. This shift of the line was first pointed out by Dirac (1927) and Oppenheimer (1930).

To give agreement with experiment the shift is, of course, neglected. It is, however, of some interest to note that it is connected with the conservation of energy. The usual expression for the intensity emitted in the frequency range, ν, $\nu+d\nu$ is

$$I(\nu)\, d\nu = \frac{\Gamma}{2\pi} \frac{\boldsymbol{h}\nu\, d\nu}{(\nu-\nu_0)^2+\Gamma^2/4},$$

where ν_0 is the resonant frequency. Clearly $\int_0^\infty I(\nu)\, d\nu \to \infty$. Further, this expression gives too low a value for frequencies greater than ν_0 (due to neglect of the varying weight factor) and even if the energy of the emitted photons is cut off at ϵ' the energy discrepancy is considerable.*

The excited state of the atom and the states containing photons do not enter the fundamental equation (11) in a symmetrical fashion. The former interacts with all the latter, while each one of the latter states only interacts with one state—the excited state of the atom. This cause one root Λ_0 of the equation (14) to be a very large negative number, whereas the other roots are close to the unperturbed energies of the states containing photons. When the energy of the initial excited states is thus depressed, a shift in the position of the line is to be expected.

It is possible to extend the set (11) so that the equations become more "symmetrical". Consider briefly the interaction between the following states:—

ψ_0—the atom in the excited state, no photons;

ψ_{1i}—the atom in the normal state, with a photon of energy ϵ_i present;

ψ_{0ij}—the atom in the excited state with two photons of energy ϵ_i, ϵ_j present.

These states do not interact with the states ψ_1, ψ_{0i}, ψ_{1ij}, if the dipole interaction alone is considered. The fundamental equations then become (with an obvious notation)

$$\left.\begin{aligned}&(E_0-\Lambda)A_0+\sum_i H_{0,1i}A_{1i}=0,\\ &(E_{1j}-\Lambda)A_{1j}+H_{1j,0}A_0+\sum_i H_{1j\,.0ij}A_{0ij}=0,\\ &(E_{0i_j}-\Lambda)A_{0ij}+H_{0ij,1i}A_{1j}+H_{0ij\,1}A_{1j}=0.\end{aligned}\right\}\quad\ldots\ldots(28)$$

---

* With the formula above

$$\int^{c/a_0} I(\nu)\, d\nu = \boldsymbol{h}\nu_0 + 1{\cdot}2\, \Gamma\boldsymbol{h}.$$

On eliminating the amplitudes A_{0ij}, and writing $H_{0ij,1j}=H_{0j,1}$, etc., equations (28) become

$$\left.\begin{aligned}\left\{E_{1j}-\Lambda-\sum_i\frac{|H_{1,0i}|^2}{E_{0ij}-\Lambda}\right\}A_{1j}-\sum_i\frac{H_{1j,0}\,H_{0\,1i}}{E_0-\Lambda}A_{1i}-\sum_i\frac{H_{1,0i}\,H_{0j,1}}{E_{0ij}-\Lambda}A_{1i}=0\\ (E_0-\Lambda)A_0+\sum_i H_{0,1i}A_{1i}=0.\end{aligned}\right\}\quad\ldots\ldots(29)$$

The original set (11) can be written in a similar form

$$\left.\begin{aligned}(E_{1j}-\Lambda)A_{1j}-\sum_i\frac{H_{1j,0}\,H_{0,1i}}{E_0-\Lambda}A_{1i}=0,\\ (E_0-\Lambda)A_0+\sum_i H_{0,1i}\,A_{1i}=0.\end{aligned}\right\}\quad\ldots\ldots(30)$$

If Λ is much less than the range of energy values E_{0ij} $(i=1, 2, \ldots .\ M)$ the last term in the first equation of (29) can be neglected compared with the preceding term. Equations (30) lead to a solution in which the centre of the line, and therefore the most important values of Λ, are given by

$$\Lambda\simeq E_0+R(E_0)=E_0+\int_{(E_0)}(E_0-E_{1i})^{-1}\,.\int|H_{0,1i}|^2\,d\Omega\,.\,\rho(E)_{1i}\,dE_{1i}$$

(neglecting the variation of $R(E)$ with E). Thus (29) will lead to a solution in which the centre of the line is given by

$$E_{1j}+\sum_i\frac{|H_{0\,1i}|^2}{E_{1j}-E_{0ij}}\simeq E_0+R(E_0).\qquad\ldots\ldots(31)$$

The difference between the most important E_{1j}, and E_0 is now reduced to a term of the form $\sum_i\frac{|H_{0,1i}|^2}{(E_0-E_{1i})^2}$. This suggests that the divergence difficulties in the emission case *may* be eliminated by taking account of higher-order processes which appear to have no physical importance. In equations (28) terms have been added which allow the atom to jump from the normal state to the excited state *emitting* a photon, and to jump back to the normal state *absorbing* a photon. Presumably the last term in the first of equations (29) modifies the shape of the line so as to reduce the energy discrepancy which arises when the shift is neglected.

§6. THE SOLUTION OF THE SCATTERING PROBLEM

The fundamental equations are

$$(E_s-\Lambda_t)A_s^t+\sum_{r}{}_{(s)}H_{sr}A_r^t=0\quad(s, t=0, 1, 2, \ldots\ M).\qquad\ldots\ldots(18)$$

Substituting $(\Lambda_t-E_s)A_s^t=V_{st}$ gives

$$V_{st}=\sum_r\frac{H_{sr}\,V_{rt}}{\Lambda_t-E_r}\qquad\ldots\ldots(32)$$

(where $H_{ss}=0$) provided $\Lambda_t\neq E_r$ for any t, r. It will be seen that all the $(M+1)$ roots Λ_t can be found in spite of this latter assumption. The r and s in (32) must include all variables required to describe the states (i.e. spins and polarizations as well as the momentum and energy).

The variation of H_{rs} with the energy of those states which are of interest will be small, and it can be neglected, if only those final states which lie in a small energy region are considered. Then, if l_1 and l_2 denote states with parallel directions * and different energies,

$$V_{l_1 t}=\sum_r \frac{H_{l_1 r} V_{rt}}{\Lambda_t - E_r}, \qquad V_{l_2 t}=\sum_r \frac{H_{l_2 r} V_{rt}}{\Lambda_t - E_r}\ ;$$

so $V_{l_1 t}=V_{l_2 t}$, i.e. the function V_{ls} depends only on the "direction" of ψ_l and is independent of E_l. Thus (32) becomes

$$V_{lt}=\sum_{E_r} \frac{1}{\Lambda_t - E_r} \sum_{[r]} H_{lr} V_{rt}, \qquad \ldots\ldots(33)$$

where $\sum_{[r]}$ denotes the sum over all states ψ_r belonging to any one energy level, and $\sum_{E_r}$ denotes the sum over the distinct energy levels. Equation (33) separates into an angular and an energy equation.

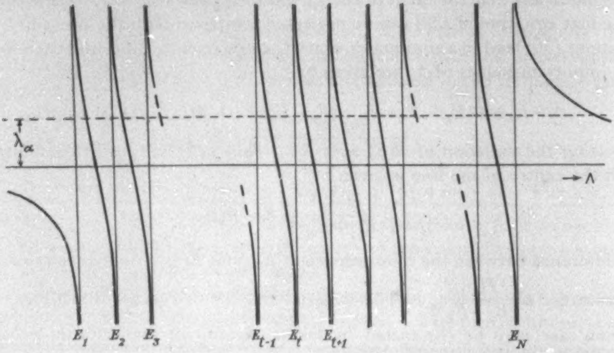


Figure 2. The intersections of the line λ_α and the curve $\sum_{E_t} \frac{1}{\Lambda - E_t}$.

Solutions of the equation

$$V_{l\alpha}=\lambda_\alpha \cdot \sum_{r=1}^{m.\eta} H_{lr} V_{r\alpha}, \qquad \ldots\ldots(34)$$

where m is the number of momentum values for each energy value, η is the spin and polarization degeneracy of each state, and λ_α is a real eigenvalue, are solutions of (33). The roots Λ_t are determined from (33) by

$$\lambda_\alpha=\sum_{E_r} \frac{1}{\Lambda_t - E_r}. \qquad \ldots\ldots(35)$$

Figure 2 indicates the graphical solution of equation (35). Between each adjacent pair of energy levels E_r, E_{r+1}, there are as many roots Λ_t as there are finite eigenvalues λ_α. If λ_α is infinite, i.e. if solutions of the equation $\sum_{r=1}^{m.\eta} H_{lr} V_{r\alpha}=0$ exist; the corresponding Λ are identical with the energy levels E_r. A small λ_α ($\lambda_\alpha \simeq 1/\Delta E$ where $\Delta E = E_{r+1} - E_r$) leads to roots which lie roughly midway between the E_r.

---

* "Directions" must include spin and polarization variables.

If there are N distinct energy levels E_r, then $m \,.\, \eta \,.\, N = M+1$ (the total number of states) and as equation (34) in general has $m\eta$ eigenvalues λ_α, there are in general $M+1$ roots Λ_t. The last $m\eta$ roots of (35) are less than the lowest E_r or greater than the highest one.

It can be assumed that the m directions $[r]$ are uniformly distributed over the solid angle 4π, so that there are $(m/4\pi) \,.\, d\Omega_r$ states in the element of solid angle $d\Omega_r$. If m is very large, equation (34) can, with little error, be written in the form

$$V'_{l_\zeta \alpha} = \lambda_\alpha \,.\, \frac{m}{4\pi} \sum_{\xi=1}^{\eta} \int H_{l_\zeta r_\xi} V'_{r_\xi \alpha} d\Omega_r, \qquad \ldots\ldots (36)$$

where ψ_{r_ξ}, ψ_{r_ζ} are the various states with momentum r, and different spins or polarizations.

If $\eta \,.\, \rho(E) \,.\, dE \,.\, d\Omega_r$ is the number of states in the energy range E, $E+dE$ whose momenta lie in the solid angle $d\Omega_r$, then $\rho(E)\Delta E = m/4\pi$. Further, if $H'_{l_\zeta r_\xi}$ is the value of the matrix element for a box of volume unity, and $\rho'(E)$ is the corresponding density function, then $\rho(E) H_{l_\zeta r_\xi} = \rho'(E) H'_{l_\zeta r_\xi}$ and equation (36) becomes

$$V'_{l_\zeta \alpha} = (\lambda_\alpha \Delta E) \rho'(E) \sum_{\xi=1}^{\eta} \int H'_{l_\zeta r_\xi} V'_{r_\xi \alpha} \, d\Omega_r. \qquad \ldots\ldots (37)$$

Thus $(\lambda_\alpha \,.\, \Delta E)$ is independent of ΔE.

It can be readily seen that for the non-relativistic scattering of photons by an electron

$$\lambda_\alpha \,.\, \Delta E = O(1) \,.\, \frac{8\pi^2}{1+\mu^2/E^2} \,.\, (\hbar c/e^2),$$

where E is the total energy, and $O(1)$ denotes a constant of the order of magnitude of unity. Thus for weak coupling between the field and the scatterer $|\lambda \, \Delta E| \geqslant 1$. It is the contrary case of strong coupling when $|\lambda_\alpha \Delta E| = O(1)$, which shows important differences between the usual approximate solution and the exact solution.

It will be sufficient to show that the usual perturbation method (even allowing for degeneracy) fails when applied to the fundamental equations (18) unless $|\lambda_\alpha \Delta E| \geqslant 1$. A matrix $\boldsymbol{S}$ is required such that

$$\boldsymbol{S}^{-1} \,.\, \boldsymbol{K} \,.\, \boldsymbol{S} = \boldsymbol{W},$$

where $K_{rs} = E_r \delta_{rs} + H_{rs}$, and $\boldsymbol{W}$ is to be diagonal. Writing $H_{rs} = e^2 H^{11}_{rs}$ where e^2 is a small parameter, an approximate solution has to be found of the form $S = S_0 + e^2 S_1 + e^4 S_2 + \ldots$, $W = W_0 + e^2 W_1 + e^4 W_2 + \ldots$. The zero approximation is $H_0 S_0 = S_0 W_0$, where $(H_0)_{rs} = E_r \delta_{rs}$, and $W_0 = H_0$. This merely shows that S_0 reduces into matrices corresponding to the states in each distinct energy level. The first-order approximation is $H_0 S_1 - S_1 H_0 + H^{11} S_0 = S_0 W_1$. Denoting states belonging to the energy level E_a by the suffix a this gives

$$\sum_{p_a} H_{n_a p_a} \,.\, S_0(p_a m_a) = S_0(n_a m_a) \,.\, (W_1)_{m_a} \,.\, e^2 \qquad \ldots\ldots (38)$$

Equations (38) and (34) are identical so $S_0(p_a m_a) = V_{p, m_a}$ and $e^2 (W_1)_{m_a} = 1/\lambda_{m_a}$. Thus

$$e^2 (W_1)_{m_a} / \Delta E = 1/(\lambda_{m_a} \Delta E). \qquad \ldots\ldots (39)$$

It is clear from consideration of equation (35) or figure 2 that $e^2 (W_1)_{m_a}$ only gives a reasonable approximation to the actual energy perturbation when $|\lambda_{m_a} \,.\, \Delta E| \geqslant 1$. Otherwise (39) gives a quite incorrect result.

§7. DEGENERACY AND NORMALIZATION

The basic integral equation (37) may be degenerate. For example, if $H'_{lr}=H$, a constant, then there is only one solution $V_{r_1}=\text{const.}$, $\lambda_1 \Delta E = 1/H \,.\, \rho'(E) \,.\, 4\pi$; so there is an infinity of linearly independent degenerate solutions, with infinite eigenvalues, satisfying $\int V_{r\alpha} d\Omega_r = 0$. In this case an infinity of roots satisfy $\Lambda = E_t$. However, this type of degeneracy is trivial and can be easily removed.

Normalize the solutions of (37) so that

$$\sum_{\xi=1}^{\eta} \int \overline{V}'_{l_\xi \alpha} V_{r_\xi \beta} d\Omega_r = \delta_{\alpha\beta}. \qquad \ldots\ldots(40)$$

Then by a well known theorem H'_{lr} can be expressed in the form

$$\rho'(E) H'_{l_\xi r_\xi} = \sum_{\alpha=1}^{\infty} V'_{l_\xi \alpha} \overline{V}'_{r_\xi \alpha} / (\lambda_\alpha \Delta E), \qquad \ldots\ldots(41)$$

provided the series is uniformly convergent. Conversely if the $V'_{l_\xi \alpha}$ are independent normal orthogonal functions, in the sense of (40), the solutions of the integral equations (37) in which the kernel is given by (41), are the $V'_{l_\xi \alpha}$, and the corresponding eigenvalues are the $\lambda_\alpha \,.\, \Delta E$.

Assume for simplicity that in the degenerate case above there is no spin or polarization degeneracy. Replace the interaction H'_{lr} by

$$\rho'(E) H'_{lr} = \rho'(E) H + \sum_{i=1}^{\infty} Y_i(l) \overline{Y}_i(r) / b_i,$$

where $Y_i(l)$ is the ith spherical harmonic, and the b_i are a series of constants which increase in magnitude very rapidly as i increases, and b_1 is as large as desired, Then the difference between H'_{lr} and H can be made extremely small, and cannot affect the physical problem appreciably. The eigenvalues $\lambda_\alpha \Delta E$ of the integral equation will then all (except the first) be very large and so will correspond to extremely weak coupling. (It will also be seen from the final stages that the influence of the b_i is negligible.)

It is also apparent, by similar reasoning, that the transition from the linear equations (34) to the integral equation (37) is valid, provided that the change in H_{lr}, in going from any direction to any "adjacent" direction, is small compared with the value of H_{lr}.

(a) *Orthogonal properties and normalization*

The first condition (6) requires

$$\sum_{r=0}^{M} \overline{A}^s_r A^{s'}_r = \delta_{ss'};$$

or

$$\sum_{r=0}^{M} \frac{\overline{V}_{rs} V_{rs'}}{(\Lambda_s - E_r)(\Lambda_{s'} - E_r)} = \delta_{ss'}. \qquad \ldots\ldots(42)$$

Putting $s=t, \alpha$; $s'=t', \alpha'$; (41) can be written

$$\sum_{E_r} \frac{1}{(\Lambda_{t\alpha} - E_r)(\Lambda_{t'\alpha'} - E_r)} \cdot \sum_{r=1}^{m\eta} \overline{V}_{r,t\alpha} V_{r,t'\alpha'} = \delta_{tt} \,.\, \delta_{\alpha\alpha}, \qquad \ldots\ldots(42\,a)$$

where $\Lambda_{t\alpha}$ ($t=1, 2, \ldots N$) denote the roots of

$$\lambda_\alpha = \sum_{E_r} \frac{1}{\Lambda - E_r}.$$

It can easily be seen that (42 *a*) is satisfied, as follows:—

(i) from (35) it follows that

$$\sum_{E_r} \frac{1}{(\Lambda_{t\alpha}-E_r)(\Lambda_{t'\alpha}-E_r)}=0,$$

if $E_t \neq E$, i.e. if $t \neq t'$;

(ii) equation (34) has orthogonal solutions, on

$$\sum_{r=1}^{m\eta} \overline{V}_{r\alpha} V_{r\alpha'}=0, \quad \text{if } \alpha \neq \alpha'.$$

The V functions are normalized by (42), so that

$$\sum_{E_r} \frac{1}{(\Lambda_{t\alpha}-E_r)^2} \sum_{r=1}^{m\eta} |V_{r,t\alpha}|^2=1. \qquad \text{......(43)}$$

It is more useful to express (43) in terms of the solutions V' of the integral equation. To be more precise let $Z_{l_\xi\alpha}$ be the *normal* orthogonal solutions of (34) and $V'_{l_\xi\alpha}$ the normal orthogonal solutions of (37).

Then

$$\left.\begin{aligned} &\sum_{\xi=1}^{\eta} \sum_{l=1}^{m} \overline{Z}_{l_\xi\alpha} Z_{l_\xi\alpha'}=\delta_{\alpha\alpha'} \\ &\sum_{\alpha=1}^{m.\eta} \overline{Z}_{l_\xi\alpha} Z_{l'_\xi\alpha}=\delta_{ll'}.\delta_{\alpha\alpha'} \end{aligned}\right\} ; \qquad \text{......(44)}$$

and

$$\sum_{\xi=1}^{\eta} \int \overline{V}'_{l_\xi\alpha} V'_{l_\xi\alpha'}\, d\Omega_l=\delta_{\alpha\alpha'}. \qquad \text{......(45)}$$

Taking $Z_{l_\xi\alpha}=a_\alpha V_{l_\xi\alpha}$, where a_α only depends on α, the first equation of (44) gives

$$|a_\alpha|^2 \sum_{\xi=1}^{\eta} \sum_{l=1}^{m} |V'_{l_\xi\alpha}|^2=(m/4\pi) \sum_{\xi} \int |V'_{l_\xi\alpha}|^2\, d\Omega_l=1$$

so $|a_\alpha|^2=4\pi/m$.

(The second equation of (44) then takes the form

$$(4\pi/m) . \sum_{\alpha=1}^{m\eta} \overline{V}'_{l_\zeta\alpha} V'_{l'_\xi\alpha}=\delta_{ll'} . \delta_{\zeta\xi},$$

and as $m \to \infty$ this becomes

$$\lim_{m\to\infty} (4\pi/m) . \sum_{\alpha=1}^{m\eta} \overline{V}'_{l_\zeta\alpha} V'_{l'_\xi\alpha}=\delta_{ll'} . \delta_{\zeta\xi},$$

the form of the second orthogonality condition for continuous eigen-functions.)

If the $V_{l,t\alpha}$ which satisfies (43) is related to $Z_{l,\alpha}$ by

$$|V_{l,t\alpha}|^2=|b_{t\alpha}|^2 . |Z_{l,\alpha}|^2,$$

then (43) gives

$$|b_{t\alpha}|^2 \sum_{E_r} \frac{1}{(\Lambda_{t\alpha}-E_r)^2}=1, \qquad \text{......(46)}$$

determining the $|b_{t\alpha}|^2$.

Finally,

$$A^s_{r\zeta}=\frac{1}{(\Lambda_{t\alpha}-E_r)} . b_{t\alpha} . \sqrt{4\pi/m} . V'_{r_\zeta\alpha}, \qquad \text{......(47)}$$

where $s=t, \alpha$.

Before applying the series for $\cot z$ and $\operatorname{cosec}^2 z$ to the solution it is worth remembering that the number of final energy levels may be very large, but the energy range which these levels cover has to be much less than the difference between the initial energy and the nearest of the intermediate energies. The energy range of the final states cannot be very great or the method of solution used at the beginning of this section (the assumption that H_{rs} is independent of the energy) will not in general be valid. As the "line breadth" must decrease as $1/L^3$, this restriction cannot affect the physically true solution, but it may mean that some divergences which would otherwise arise are not apparent. However, allowing the breadth of the band of final energy states to increase should show up some at least of the divergences.

Equation (35) can be solved in a similar way to (14), viz.:—

$$\lambda_\alpha = \sum_{r=-n}^{+n} \frac{1}{(\Lambda_{t\alpha}-E_s)-r\Delta E} + \left(\sum_{r=1}^{s-n-1} + \sum_{r=s+n+1}^{N}\right)\frac{1}{\Lambda_{t\alpha}-E_r},$$

giving the value of $\Lambda_{t\alpha}$ in the vicinity of E_s. So

$$\lambda_\alpha = \frac{\pi}{\Delta E} \cdot \cot\{(\Lambda_{t\alpha}-E_s)\pi/\Delta E\} + \frac{1}{\Delta E}\int_{(E_s)} \frac{1}{(E_s-E_r)}\,dE_r, \qquad \ldots\ldots(48)$$

using the notation of § 5. Thus

$$\lambda_\alpha \,.\, \Delta E = \pi \cot\{(\Lambda_{t\alpha}-E_s)\pi/\Delta E\} + \sigma(E_s),$$

where

$$\sigma(E_s) = \int_{(E_s)} \frac{1}{E_s-E_r}\,dE_r\,.$$

$\sigma(E_s)$ is clearly similar to the term $R(E_s)$ of § 5. If there were no intermediate states and H_{rs} could be taken as independent of energy this term σ could become a very large negative quantity.

Equation (46) for $|b_{t\alpha}|^2$ is treated similarly to equation (21). In the vicinity of E_s

$$\sum_{E_r} \frac{1}{(\Lambda_{t\alpha}-E_r)^2} \simeq \sum_{r=-\infty}^{+\infty} \frac{1}{(\Lambda_{t\alpha}-E_s-r\Delta E)^2} = (\pi/\Delta E)^2 \operatorname{cosec}^2\{(\Lambda_{t\alpha}-E_s)\pi/\Delta E\}.$$

Here the remainder terms for large r can be neglected in a similar fashion to those arising from equation (21). Using (48) gives

$$\sum_{E_r} \frac{1}{(\Lambda_{t\alpha}-E_r)^2} = \frac{1}{(\Delta E)^2}\{\pi^2 + (\lambda_\alpha \Delta E - \sigma(E_s))^2\}$$

so

$$|b_{t\alpha}|^2 = \frac{(\Delta E)^2}{\pi^2 + (\lambda_\alpha \Delta E - \sigma(E_s))^2} \qquad \ldots\ldots(49)$$

$|b_{t\alpha}|^2$ gives the measure of the importance of the eigenvalue $(\lambda_\alpha \Delta E)$ in the solution $A_r^{t\alpha}$. $\sigma(E_s)$ must be neglected in (49) else $|\lambda_\alpha \Delta E| = O(1)$ may lead to small contributions, while $\lambda_\alpha \Delta E \gg 1$ could give a large $|b_{t\alpha}|^2$. Thus there is a divergent term* arising from the H_{rs}—the second order matrix elements in equation (18). Putting $\sigma(E_s)=0$ gives†

$$|b_{t\alpha}|^2 = \frac{(\Delta E)^2}{\pi^2 + (\lambda_\alpha \Delta E)^2}\,. \qquad \ldots\ldots(50)$$

---

* This divergence difficulty cannot be removed by the present methods.

† Equations (47) and (50) show that the "line breadth" behaves as $1/(\Lambda_s - E)^2$, so there is no breadth in the sense of the emission case.

(b) *The variation of the initial state*

If the initial state is ψ_{k_ξ} the initial conditions are

$$\left.\begin{aligned} a_{k_\xi}(0) &= 1, \\ a_s(0) &= 0 \qquad s \neq k_\xi, \end{aligned}\right\} \qquad \ldots\ldots(51)$$

so $c_\mu = \bar{A}^\mu_{k_\xi}$, and for any state ψ_{r_ξ} the time variation is given by

$$\begin{aligned} a_{r_\zeta} e^{-iE_r t} &= \sum_{\mu=0}^{M} \bar{A}^\mu_{k_\xi} A^\mu_{r_\zeta} e^{-i\Lambda_\mu t} \\ &= \sum_{\alpha=1}^{m.\eta} \sum_{p=1}^{N} \bar{A}^{p,\alpha}_{k_\xi} A^{p,\alpha}_{r} e^{-i\Lambda_{p\alpha}.t} \qquad (\mu = p, \alpha). \end{aligned}$$

In particular

$$a_{k_\xi} e^{-iE_k t} = \sum_{\alpha=1}^{m.\eta} \sum_{p=1}^{N} |A^{p,\alpha}_{k_\xi}|^2 e^{-i\Lambda_{p\alpha}.t},$$

so

$$a_{k_\zeta} = \sum_\alpha \sum_p |A^{p,\alpha}_{k_\xi}|^2 . e^{-i(n\Delta E + \delta_\alpha)t},$$

where $\Lambda_{p\alpha} - E_k = n\Delta E + \delta_\alpha$. Equation (49) states

$$(\pi/\Delta E) \cot(\pi\delta_\alpha/\Delta E) = \lambda_\alpha,$$

thus, using (47)

$$a_{k_\xi} \simeq \frac{4\pi}{m} \sum_\alpha \frac{|V'_{k_\xi\alpha}|^2}{\pi^2 \operatorname{cosec}^2(\pi\delta_\alpha/\Delta E)} . (\Delta E)^2 \sum_{n=-\infty}^{+\infty} \frac{1}{(n\Delta E + \delta_\alpha)} e^{-i(n\Delta E + \delta_\alpha)t} \ldots\ldots(52)$$

Substituting $t . \Delta E = x$, the last summation becomes

$$\frac{1}{(\Delta E)^2} \sum_{n=-\infty}^{+\infty} \frac{1}{(n + \delta_\alpha/\Delta E)^2} . e^{-i(n+\delta_\alpha/\Delta E).x}.$$

This sum is periodic in x with period 2π. In Appendix I it is evaluated. The sum is a continuous function of x, but the first derivative of the sum is discontinuous at $x = s\pi$, $s = 0, \pm 1, \pm 2, \ldots$ For $0 \leqslant x \leqslant \pi$ (52) gives

$$a_{k_\xi} = \frac{4\pi}{m} \sum_{\alpha=1}^{m.\eta} |V'_{k_\xi\alpha}|^2 \left\{1 - \pi x \frac{1 - i\cot(\pi\delta_\alpha/\Delta E)}{\pi^2 \operatorname{cosec}^2(\pi\delta_\alpha/\Delta E)}\right\}. \qquad \ldots\ldots(53)$$

When $|a_k|^2$ is evaluated the imaginary terms in (53) will give rise to terms in x^2 only. If $x \ll 2\pi$, i.e. if $t \ll 2\pi/\Delta E$, the x^2 terms can be neglected. As $|a_{k_\xi}|^2$ is only to be evaluated correctly for the constant terms and those depending on x, it is sufficient to take

$$a_{k_\xi} = \frac{4\pi}{m} \sum_{\alpha=1}^{m.\eta} |V'_{k_\xi\alpha}|^2 \left\{1 - \frac{\pi x}{\pi^2 \operatorname{cosec}^2(\pi\delta_\alpha/\Delta E)}\right\}. \qquad \ldots\ldots(54)$$

Thus the transition probability Γ is given by

$$\Gamma/2 = \Delta E . \frac{4\pi}{m} . \sum_\alpha |V_{k_\xi\alpha}|^2 \frac{\pi}{\pi^2 \operatorname{cosec}^2(\pi\delta_\alpha/\Delta E)},$$

or

$$\rho(E) . \Gamma/2 = \sum_\alpha |V'_{k_\xi\alpha}|^2 . \frac{\pi}{\pi^2 + (\lambda_\alpha \Delta E)^2}. \qquad \ldots\ldots(55)$$

Remembering that the V'_{kx} are normalized functions it is obvious that in general the higher eigenvalues $\lambda_\alpha \Delta E$ give very little contribution to the transition probability, so in practice only a few of the lowest solutions of (37) need be considered.

(c) *The scattered states*

The analogue of (52) for any state a_r $(r \neq k)$ is

$$a_{r_\zeta} = \frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} \; \frac{1}{\pi^2 \operatorname{cosec}^2(\pi\delta_\alpha/\Delta E)}$$

$$\times \sum_{n=-\infty}^{+\infty} \frac{1}{(n+q+\delta_\alpha/\Delta E)(n+\delta_\alpha/\Delta E)} \cdot e^{-i(n+q+\delta_\alpha/\Delta E).x}$$

where $\Lambda_{p\alpha} - E_r = (\Lambda_{p\alpha} - E_k) + (E_k - E_r) = (n\Delta E + \delta_\alpha) + q\Delta E,\; E_k - E_r = q\Delta E.$
In the Appendix it is shown that if $q \neq 0$

$$\sum_{n=-\infty}^{+\infty} \frac{1}{(n+q+\delta_\alpha/\Delta E)(n+\delta_\alpha/\Delta E)} \cdot e^{-i(n+q+\delta_\alpha/\Delta E).x} = \frac{\pi}{q}(e^{-iqx}-1)\{1+\cot(\pi\delta_\alpha/\Delta E)\}$$

so if $q \neq 0$ (i.e. $E_r \neq E_k$),

$$a_{r_\zeta} = \frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} \cdot \frac{1}{\pi\{-i+\cot(\pi\delta_\alpha/\Delta E)\}} \cdot \frac{1}{q}(e^{-iqx}-1);$$

and if $E_r = E_k$,

$$a_{r_\zeta} = \frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} \cdot \frac{i}{\pi\{-i+\cot(\pi\delta_\alpha/\Delta E)\}} \cdot x.$$

as

$$\frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} = 0, \quad \text{if } r \neq k.$$

Only the total probability of scattering in a given "direction" is of interest, so the probabilities for all the states of different energies having this "direction" must be added. This gives

$$\sum_{E_r} |a_r|^2 = \left| \frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} \frac{i}{\pi\{-i+\cot(\pi\delta_\alpha/\Delta E)\}} \right|^2 \cdot \left\{ x^2 + \sum_q{}' \frac{1}{q^2} |e^{-iqx}-1|^2 \right\},$$

where $\sum_q{}'$ denotes summation over all integers except zero. Further,

$$x^2 + \sum_q{}' \frac{1}{q^2} |e^{-iqx}-1|^2 = x^2 + 2\sum_q{}' \frac{1-\cos qx}{q^2} = 2\pi x$$

(see Appendix). Thus

$$\sum_{E_r} |a_{r_\zeta}|^2 = \left| \frac{4\pi}{m} \sum_\alpha \overline{V}'_{k_\xi\alpha} V'_{r_\zeta\alpha} \frac{1}{\pi + i(\lambda_\alpha \Delta E)} \right|^2 . \, 2\pi \, . \, \Delta E \, . \, t. \quad \ldots\ldots(56)$$

If $\sigma(E_s)$ were not neglected in (50) the transition probability would become

$$\rho(E) \, . \, \Gamma/2 = \sum_\alpha | V'_{k_\xi\alpha}|^2 \frac{\pi}{\pi^2 + \{\lambda_\alpha \Delta E - \sigma(E)\}^2} \; ;$$

so, except in the region of initial energy E where $\sigma(E) \simeq 0$, the physically unimportant eigenvalues satisfying $|\lambda_\alpha \Delta E| \gg 1$ might give much greater contributions than eigenvalues satisfying $|\lambda_\alpha \Delta E| = O(1)$.

In the evaluation of the series $\sum_{E_r} \frac{1}{(\Lambda_{t\alpha} - E_r)^2}$, which gave $|b_{t\alpha}|^2$, it is necessary (as in the emission case) to assume that the summation is cut off at some arbitrarily large energy value,

§8. THE HEITLER-PENG INTEGRAL EQUATION

The Heitler-Peng integral equation

$$U_{r_\zeta k_\xi}=H'_{r_\zeta k_\xi}+i\pi\rho'(E)\sum_{\xi'=1}^{\eta}\int H'_{r_\zeta l_{\xi'}}U_{l_{\xi'}k_\xi}\,d\Omega_l, \qquad \ldots\ldots(57)$$

now appears as a useful method for making the summations occurring in the expressions (55) and (56).

From the homogeneous linear integral equation (37), whose solutions satisfy the normalizing conditions (40), it is possible to construct a non-homogeneous linear integral equation

$$U_{r_\zeta}=F_{r_\zeta}+g\rho'(E)\sum_{\xi'=1}^{\eta}\int H'_{r_\zeta l_{\xi'}}U_{l_{\xi'}}\,d\Omega_l, \qquad \ldots\ldots(58)$$

where g is, at present, an unknown constant, and F_{r_ζ} an unkown function of r_ζ.

The solution U_{r_ζ} of (58) can be made to give the $\sum\limits_\alpha$ term in (56).

If $U_{r_\zeta}=\sum\limits_\alpha u_\alpha V'_{r_\zeta\alpha}$ $F_{r_\zeta}=\sum\limits_\alpha f_\alpha V'_{r_\zeta\alpha}$, the solution of (58) is given by

$$u_\alpha=\mu_\alpha f_\alpha/(\mu_\alpha-g), \quad \text{where } \mu_\alpha=\lambda_\alpha\Delta E.$$

Putting $\rho'(E)\mu_\alpha f_\alpha=\overline{V}'_{k_\xi\alpha}$, $g=i\pi$ gives

$$\rho'(E)\,U_{r_\zeta}=\rho'(E)\sum_\alpha u_\alpha V'_{r_\zeta\alpha}=i\sum_\alpha\frac{\overline{V}'_{k_\xi\alpha}V'_{r_\zeta\alpha}}{\pi+i(\lambda_\alpha\Delta E)}. \qquad \ldots\ldots(59)$$

Writing $U_{r_\zeta}=U_{r_\zeta k_\xi}$, the equation (56) becomes

$$\sum_{Er}|a_r|^2=2\pi x\,.\left|\frac{4\pi}{m}\rho'(E)\,U_{r_\zeta k_\xi}\right|^2.$$

The probability for scattering in the element of solid angle $d\Omega_r$ about r is given by

$$\rho(E)\,.\,\Delta E\,.\,d\Omega_r\sum_{Er}|a_{r_\zeta}|^2=2\pi t\,.\,\rho(E)\,.\,(\Delta E)^2\left|\frac{4\pi}{m}\rho'(E)\,U_{r_\zeta k_\xi}\right|^2.\,d\Omega_r.$$

Using the relation* $(4\pi/m)\,.\,\rho'(E)=(L^3\,.\,\Delta E)^{-1}$, the transition probability γ for scattering into the solid angle $d\Omega_r$ becomes

$$\gamma=\frac{1}{L^3}\,.\,2\pi\rho'(E)\,|U_{r_\zeta k_\xi}|^2\,.\,d\Omega_r \quad (\hbar=1), \qquad \ldots\ldots(60)$$

where $U_{r_\zeta k_\xi}$ satisfies equation (57).

The Heitler-Peng result for the transition probability of the initial state

$$\Gamma=\frac{1}{L^3}\,.\,2\pi\,\mathscr{R}\left[\rho'(E)\sum_{\zeta=1}^{\eta}\int H'_{k_\xi r_\zeta}U_{r_\zeta k_\xi}\,d\Omega_r\right] \qquad \ldots\ldots(61)$$

can easily be verified.

$$\rho'(E)\sum_{\zeta=1}^{\eta}\int H'_{k_\xi r_\zeta}U_{r_\zeta k_\xi}\,d\Omega_r=\frac{1}{\rho'(E)}\,.\sum_{\zeta=1}^{\eta}\int\sum_\alpha\sum_\beta\frac{V'_{k_\xi\alpha}\overline{V}_{r_\zeta\beta}}{\mu_\alpha}\,.\,\frac{V'_{r_\zeta\beta}\overline{V}_{k_\xi\beta}}{\mu_\beta-i\pi}\,.\,d\Omega_r$$

using (59) and the expransion of H'_{kr} in terms of eigensolutions. Thus

$$\rho'(E)\sum_{\zeta=1}^{\eta}\int H'_{\xi r_\zeta}U_{r_\zeta k_\xi}\,d\Omega_r=\frac{1}{\rho'(E)}\left\{\sum_\alpha\frac{|V'_{k_\xi\alpha}|^2}{\mu_\alpha+\pi^2}+i\pi\sum_\alpha\frac{|V_{k_\xi\alpha}|^2}{\mu_\alpha(\mu_\alpha^2+\pi^2)}\right\},$$

---

* As above, $\rho(E)$ and H_{rt} are the density function and matrix elements for a cubic box with edges of length L, while $\rho'(E)$ and H'_{lt} are the same quantities for a cubic box with unit edges.

so (61) and (55) are identical. It can, however, be seen that a simpler expression would be

$$\Gamma = \frac{1}{L^3} \cdot 2\pi\rho'(E) \sum_{\zeta=1}^{\eta} \int |U_{r\zeta k_\xi}|^2 \, d\Omega_r, \qquad \ldots\ldots(62)$$

which is also identical with (55). Those transitions which lead from the initial state ψ_{k_ξ} to other states with the same direction but different energies can be entirely neglected.

Gormley and Heitler (1944) have shown that the equations (60) and (61) have the correct Lorentz transformation properties. Thus (55) and (56) will give the transition probabilities when the total momentum of the system is not zero, and in cases in which the energy levels are no longer almost equi-distant, and almost equally degenerate.* In a new Lorentz frame of reference, H_{rs}, the interaction matrix, will have a different form, and consequently the λ_α and $V'_{r\alpha}$ will change in going from one frame of reference to another. It may be of advantage to choose the frame of reference in which H_{rs} is least dependent on the variables, as equation (37) can possibly be solved by approximate methods in that case.

Finally, it is worth noticing that equation (41) assumes that the correct normalization condition is $\sum_{r=0}^{M} \bar{A}_r^s A_r^{s'} = \delta_{ss'}$; while the correct condition is that the total probability of the system being in any of the final *or intermediate* states is unity. The treatment given in §§ 6 and 7 is thus only valid provided $\sum_i |A_i|^2 \ll 1$, where A_i is the amplitude of an intermediate state.

Using the second of equations (15) and the values of A_r given in § 7 it can be shown that this condition holds provided the initial energy lies close to the centre of the energy range of the final states. Otherwise it is not true; and the intermediate states thus give rise to a further divergence difficulty in the solution.

ACKNOWLEDGMENTS

I am indebted to Professor Dirac, and to Drs. B. Ferretti and H. W. Peng, for useful discussion.

APPENDIX

Various summations

1. The values of

$$\left.\begin{aligned} f(x) &= \sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{n+a} \\ g(x) &= \sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{(n+a)^2} \end{aligned}\right\} \quad (a \neq \text{an integer})$$

are required.

Now

$$\sum_{r=-n}^{+n} \frac{e^{irx}}{r+a} \cdot e^{iax} = i \int_0^x e^{iat} \sum_{r=-n}^{+n} e^{irt} . \, dt + \sum_{r=-n}^{+n} \frac{1}{r+a},$$

and

$$\sum_{r=-n}^{+n} e^{irt} = \frac{\sin (n+\frac{1}{2})t}{\sin (t/2)}.$$

---

* If $\rho'(E)$ depends on the angles it must be included in the kernels of the integral equations.

Further, if $0 \leqslant x \leqslant \pi$,

$$\int_0^x \frac{\sin(n+\frac{1}{2})t}{\sin(t/2)}\, e^{iat}\, dt = 2\int_0^x e^{iat}\frac{\sin(n+\frac{1}{2})t}{t}\, dt + \int_0^x e^{iat}\sin(n+\tfrac{1}{2})t\left\{\frac{1}{\sin(t/2)} - \frac{1}{t/2}\right\}dt.$$

The last term equals

$$\left[e^{iat}\left(\frac{1}{\sin(t/2)} - \frac{1}{t/2}\right)\frac{\cos(n+\frac{1}{2})t}{n+\frac{1}{2}}\right]_0^x$$

$$+ \frac{1}{n+\frac{1}{2}}\int_0^x \cos(n+\tfrac{1}{2})t \,.\, \frac{d}{dt}\left\{e^{iat}\left(\frac{1}{\sin(t/2)} - \frac{1}{t/2}\right)\right\}dt$$

and as $\dfrac{1}{\sin y} - \dfrac{1}{y}$, and $\dfrac{d}{dy}\left(\dfrac{1}{\sin y} - \dfrac{1}{y}\right)$ are bounded in the range $0 \leqslant y \leqslant \pi/2$, these terms are $O(1/n)$. Thus

$$\sum_{r=-\infty}^{+\infty} \frac{e^{irx}}{r+a} \,.\, e^{iax} = \pi\{i + \cot a\pi\}$$

as

$$\sum_{r=-\infty}^{+\infty} \frac{1}{r+a} = \pi \cot a\pi.$$

Thus

$$f(x)e^{iax} = \pi\{i + \cot a\pi\} \qquad (0 \leqslant x \leqslant \pi)$$

and

$$= \pi\{-i + \cot a\pi\} \qquad (-\pi \leqslant x \leqslant 0).$$

The function $f(x)$ obviously has period 2π.

Similarly, if $0 \leqslant x \leqslant \pi$,

$$\sum_{r=-n}^{+n} \frac{e^{irx}}{(r+a)^2} \,.\, e^{iax} = i\int_0^x \sum_{r=-n}^{+n} \frac{e^{irt}}{r+a} \,.\, e^{iat} \,.\, dt + \sum_{r=-n}^{+n} \frac{1}{(r+a)^2}.$$

The series for $f(x)e^{iax}$ is uniformly convergent over $(0, \pi)$ except in the neighbourhood of 0 and π, and it is boundedly convergent over the whole interval, so we may write

$$\sum_{r=-\infty}^{+\infty} \frac{e^{irx}}{(r+a)^2} \,.\, e^{iax} = i\int_0^x \pi(i + \cot a\pi)dt + \sum_{r=-\infty}^{+\infty} \frac{1}{(r+a)^2} \qquad (0 \leqslant x \leqslant \pi).$$

Thus

$$g(x)e^{iax} = i\pi(i + \cot a\pi)x + \pi^2 \operatorname{cosec}^2 a\pi \qquad (0 \leqslant x \leqslant \pi)$$

$$= i\pi(-i + \cot a\pi)x + \pi^2 \operatorname{cosec}^2 a\pi \qquad (-\pi \leqslant x \leqslant 0).$$

2. The value of

$$\sideset{}{'}\sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{n^2} \qquad (0 \leqslant x \leqslant \pi).$$

As before

$$\sum_{r=-\infty}^{+\infty} \frac{e^{irx}}{r} = \lim_{n\to\infty} i\int_0^x \sideset{}{'}\sum_{r=-n}^{+n} e^{irt} \,.\, dt + \sideset{}{'}\sum_{r=-\infty}^{+\infty} \frac{1}{r}$$

$$= \lim_{n\to\infty} i\int_0^x \left\{\frac{\sin(n+\frac{1}{2})t}{\sin(t/2)} - 1\right\}dt$$

$$= i(\pi - x) \qquad (0 \leqslant x \leqslant \pi).$$

Thus

$$\sideset{}{'}\sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{n^2} = i\int_0^x i(\pi - x)dx + \sideset{}{'}\sum_{n=-\infty}^{+\infty} \frac{1}{n^2}$$

$$= -\pi x + \frac{x^2}{2} + \frac{\pi^2}{3}.$$

Further

$$\sum_{r=-\infty}^{+\infty}{}' \frac{1-\cos rx}{r^2} = \sum_r{}' \frac{1}{r^2} - \sum_r{}' \frac{e^{irx}}{r^2}$$

$$= \pi x - \frac{x^2}{2} \qquad (0 \leqslant x \leqslant \pi).$$

3. The sum

$$\sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{(n+a)(n+b)} \qquad \text{if } a=m+b,$$

where m is a non-zero integer.

If $0 \leqslant x \leqslant \pi$, we have

$$\sum_{r=-n}^{+n} \frac{e^{irx} \,.\, e^{ibx}}{(r+a)(r+b)} = i\int_0^x \sum_{r=-n}^{+n} \frac{e^{irx}}{r+a} \,.\, e^{ibx}\,dx + \sum_{r=-n}^{+n} \frac{1}{(r+a)(r+b)}.$$

The latter sum is zero, and as before we have

$$\sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{(n+a)(n+b)} \,.\, e^{ibx} = i\int_0^x \pi(i+\cot a\pi)e^{i(b-a)t}\,dt;$$

so

$$\sum_{n=-\infty}^{+\infty} \frac{e^{inx}}{(n+a)(n+b)} = \frac{\pi}{b-a}(i+\cot a\pi)(e^{-iax}-e^{-ibx}).$$

REFERENCES

Dirac, 1927. *Z.S.*, **44**, 585.
Gormley and Heitler, 1944. *Proc. Roy. Irish Acad.*, 50 A, **4**, 29.
Heitler, 1941. *Proc. Camb. Phil. Soc.*, **37**, 291.
Heitler and Peng, 1942. *Proc. Camb. Phil. Soc.*, **38**, 296.
Oppenheimer, 1930. *Phys. Rev.*, **35**, 461.
Peng, 1944. *Nature, Lond.*, **154**, 544.
Peng, 1946. *Proc. Roy. Soc.*, A, **186**, 119.
Weisskopf and Wigner, 1930. *Z.S.*, **63**, 54 ; 1930. *Ibid.*, **65**, 18.
Wilson, 1941. *Proc. Camb. Phil. Soc.*, **37**, 301.

A WAVEFRONT SHEARING INTERFEROMETER

By W. J. BATES,
H. H. Wills Physical Laboratory, University of Bristol

MS. received 26 *December* 1946

ABSTRACT. A new type of interferometer is described by means of which the asphericity of an optical wavefront can be measured, by testing it against itself with lateral displacement or shear. Continuous control of the amount of this shear, and of the meridional or sagittal fringes is obtained in white light.

§1. INTRODUCTION

The possibility of determining uniquely the shape of a wavefront by interferometric examination on superposition with another wavefront seems not to have been exploited fully. The usual methods involve the provision of a substantially error-free wavefront as reference standard. When, however, one is allowed to displace one wavefront laterally with respect to the other, or to introduce a "shear" between them, then the error-free standard may be unnecessary; for the asphericities of two *dissimilar* wavefronts of revolution symmetry (about known centres), may be determined uniquely by a single interference pattern.

The existence of such solutions may be seen in the following way: Let W_1 and W_2 (figure 1) be two such wavefronts, giving an interferogram in their overlap region; and let the distance between the centres of revolution symmetry, C_1 and C_2, be greater than zero, and less than or equal to the distance AB. Then the asphericity of W_1 inside the circle α may be determined by the fringe intersections with the circle β around the arc EC_1F, this asphericity being referred to a sphere passing through β and C_2. The asphericity of W_2 inside the circle β may be obtained in a similar way, and the tilt and difference of curvature between these reference caps is also obtained. The process may be extended step by step along the line AB to include the whole of both wavefronts, with no assumptions as to absence of discontinuities. If the circles α and β intersect or touch the peripheries of W_1 and W_2, then the complete solution is obtained in a single step; and if C_1 and C_2 lie outside the interferogram then a solution is still possible in the interferogram region. It is clear also that when the distance C_1C_2 is zero then no solution is possible.

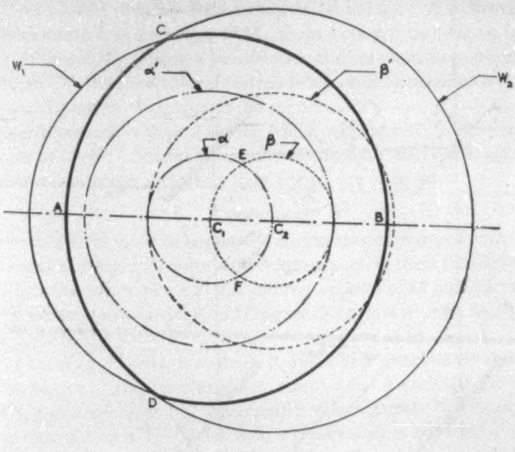


Figure 1.

This paper is concerned with the case when the two wavefronts are ***identical***, a wavefront being tested against a "sheared sight of itself". It will be shown later that the existence of solutions is not confined to the case of wavefronts possessing revolution symmetry, but may also be extended to include simple astigmatism and coma.

When one is permitted to rotate one wavefront about a principal ray, in addition to being allowed lateral shear, then added information on the asphericities may be

obtained. For example, with a relative rotation through π the comatic errors will appear doubled, and with a rotation through $\pi/2$ the astigmatic errors will appear doubled. The wavefront shearing interferometer to be described produces both lateral shear and tilt. Other interferometers have been devised which produce rotatory shear about a principal ray, in addition to lateral shear and tilt.

§ 2. PRINCIPLE OF THE INSTRUMENT

Essentially the interferometer consists of two plane dividing films, D and S, and two plane mirrors, M_L and M_R (figure 2). A convergent wavefront S_1, of

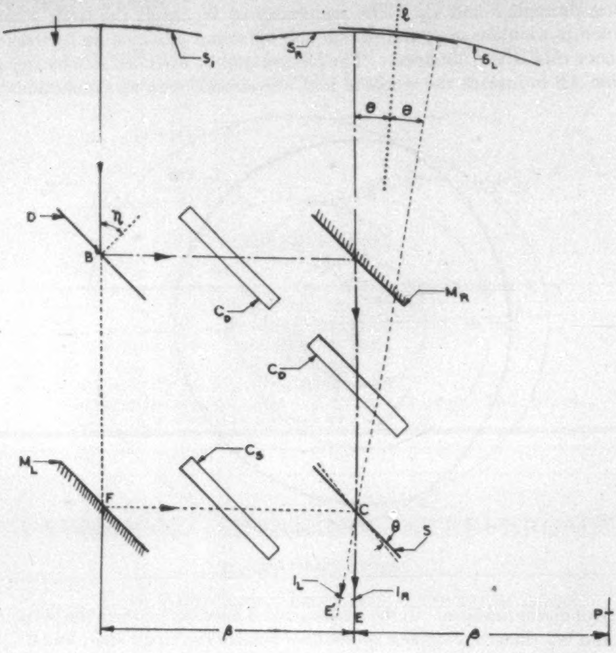


Figure 2.

which a principal ray meets D in B, is divided, and two identical coherent wavefronts emerge, one by reflexion in M_R and transmission through S (the R path) and one by reflexion in M_L and S (the L path), with principal rays along CE and CE′ respectively. An eye placed to receive these emergent wavefronts will see two apertures sheared with respect to one another (figure 2). The magnitude of the shear is continuously variable by rotation of S, the shear plate.

The number of fringes and the tilt between the wavefronts will depend on the relative positions of the two images I_L and I_R.

Separations in the plane of the paper give fringes perpendicular to that plane and path length changes, while separations out of the plane of the paper give meridional section fringes. The angle between the emergent principal rays will be twice the difference between the angles of the mirror pairs D, M_R and S, M_L. If the latter two angles are made zero, then the apparatus consists of two parallel mirror pairs, and there is no shear between the emergent wavefronts. When the shear plate S is rotated through an angle θ about any point, the angle between CE and CE′ is 2θ. If the axis of rotation is chosen so that on shearing there is no relative positional change between I_L and I_R, then the number, tilt and whiteness of the fringes, will change only because of the wavefront's relative shear. This is so when I_L and I_R are located on the shear plate at C, and the latter is rotated about an axis in its plane passing through C during the shearing operation. Equalization of the path lengths may be effected exactly by a linear translation of M_L along BF, together with a rotation about F, so that the normal to M_L always bisects the angle BFC—a motion possible with a mechanical linkage. A sufficient approximation to this control, is a rotation of M_L about a point P twice as far from BF as from CE. If D and M_L remain a parallel pair, and are rotated as a whole about an axis parallel to CE, then meridional section fringes will be obtained by the separation of I_L and I_R out of the plane of the paper.

The interferometer is then a device which will interfere a wavefront with a sheared sight of itself. *Continuous* control of the shear, and of the meridional or sagittal fringes is effected and white light fringes are obtained.

Compensation

The dividing films must be supported on transparent plates of finite thickness, and when shear is present an automatic compensation of these thicknesses may be made. When the components of the interferometer are all parallel and the test wavefront is incident at an angle η on D (figure 2), there is no shear and by both the R and the L paths the wavefront encounters one plate at an angle η. If shear is obtained by rotating S through an angle θ, and a new principal ray R be defined, through the centre of the interferogram, then by the R path the wavefront will encounter S at an angle η, and by the L path it will encounter D at an angle $(\eta+\theta)$. To effect compensation of this difference, it is necessary to add two further plates, C_D and C_S, of the same thicknesses respectively as D and S. C_S is fixed parallel to the shear plate and rotates with it during the shearing operation, whilst the second plate, if placed at C_D, is made to rotate at twice the rate of the shear plate and in the opposite direction, and if placed at C'_D at twice the rate in the same direction. Exact compensation may thus be effected by a 1 : 2 gear ratio between S and C_D, or an equivalent mechanical linkage. The errors introduced by exact compensation are investigated in the appendix.

§ 3. EXPERIMENTAL

The apparatus in an experimental form is shown in plate 2. All the plates are in holders which are kinematically attached to their supports, and all possess altazimuth line-up adjustments. The plate D and mirror M_L (a parallel pair) can be rotated as a whole by small amounts by means of the screw T, thus obtaining the vertical image separation necessary for meridian section fringes. The path length control is the approximate one mentioned in § 2, working on screw control L, and pivoting about the point P. The 1 : 2 gear ratio between the shear plate S, plus its compensator C_S, and the dividing plate compensator C_D, is anchored through a steel tape to two drums on the underside of the apparatus. The shear control is by means of the screw A working on a lever attached to the drum below C_D. Finally the apparatus is on a carriage, which can be moved in two directions at right-angles, in order to position the wavefront "focus" on the shear plate at its axis of rotation.

Departures from flatness of the plates and mirrors over the areas used, will appear as errors of the same order of magnitude in the interferogram interpretations. Relative variations in the plate thicknesses will also introduce spurious aberrations, for the magnitudes of the plate aberrations are proportional to their thicknesses, and if these are not correct then the compensation will be inexact. It is shown in the appendix that when, for instance, it is required to test an F.5 cone to an accuracy of 1/10 fringe, the plate thicknesses must be equal in pairs to about 1/30 mm. for plates 5 mm. thick. This tolerance will be proportional to the desired accuracy, and inversely proportional to the square of the test aperture. In a similar way errors in plate orientations lead to errors of compensation, and it is shown that for the same accuracy of measurement, the plate orientations must be correct to about 9′ of arc. Thus sufficiently exact compensation is not difficult.

Even so, when no compensating plates are used, the interferometer still has its uses. In this condition the path lengths in glass do not remain equal on shearing, and in addition astigmatic and comatic errors are introduced by the difference in plate orientations. Choosing the meridional focal setting, the fringes, though unequally spaced, will be straight in the absence of spherical or zonal aberration. The limit of accuracy of the instrument then depends on the spurious comatic error, and setting the maximum tolerable error at 1/100 fringe, it may be seen that it is possible in this way to test an F.7 cone sheared "edge over centre".

A restricted source is necessary and the permissible size may easily be calculated. In the direction parallel to the plates, the source may be extended provided that the slit images are parallel to the axis of rotation of the shear plate. Notwithstanding this, there is no point in extending the images beyond the eye pupil diameter for visual observation. In the direction of shear, the source size is limited by the fact that during the shearing operation the images are rotated with respect to one another. The fringe visibility in a wavelength λ, with a slit width d, and shear angle θ (figure 2) may be calculated by considering the two wavefronts which leave the aperture as having been produced by two identical sources of width d rotated through an angle ψ with respect to each other, as shown in figure 2 (a). We are concerned with the phases of the waves from the sets of points A($\equiv$A′), B($\equiv$B′), …, and the intensity at a point defined by the parameter k is

$$I_k = \int_{-d/2}^{d/2} \cos^2\left(\frac{2\pi x\psi}{\lambda} + k\right) dx$$

where x, as shown, is the distance from the intersection of the sources, and the limits of integration represent the whole width of the slit. Writing $\pi\psi/\lambda = a$ and $2ax + k = z$, the intensity becomes

$$I_k = \frac{1}{2a}\int_{k-ad}^{k+ad} \cos^2 z \,.\, dz = \frac{d}{2} + \frac{\lambda}{4\pi\psi} \cos 2k \sin \frac{2\pi\psi d}{\lambda},$$

so that the maxima and minima of I are

$$\frac{d}{2} \pm \frac{\lambda}{4\pi\psi} \sin \frac{2\pi\psi d}{\lambda}$$

and the visibility of the fringes

$$V = \frac{I_{\max} - I_{\min}}{I_{\max}} = \frac{\dfrac{\lambda}{2\pi\psi} \sin \dfrac{2\pi\psi d}{\lambda}}{\dfrac{d}{2} + \dfrac{\lambda}{4\pi\psi} \sin \dfrac{2\pi\psi d}{\lambda}}.$$

Now the shear angle θ is $\psi/2$, and if we write $4\pi\theta d/\lambda = \phi$, the expression for the visibility becomes

$$V = \frac{2 \sin \phi}{\phi + \sin \phi}.$$

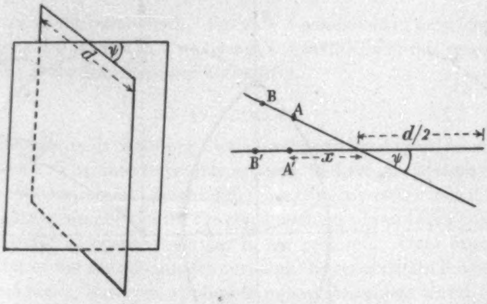


Figure 2 *a*.

For an F.10 cone sheared until the aperture edges are over the centres, and in a wavelength of 5000 A., this expression gives the permissible slit width for 60% visibility as $3{\cdot}25 \times 10^{-4}$ cm. This particular visibility is, as a matter of fact, about the value obtained with contact fringes between unsilvered plates, but much smaller values can be used satisfactorily.

The production of slits of the necessary width is not difficult, and one has been made which is adjustable from 2×10^{-2} to 1×10^{-4} inches with certainty. Diffraction around the sides of the slit causes no trouble.

§ 4. CHARACTERISTICS OF SHEARED WAVEFRONTS

When two wavefronts containing identical asphericities are exactly overlapped, a single bright fringe will cover the field of view, and if a tilt is applied between them then straight fringes parallel to the intersection of the wavefronts will result. On shearing, an interferogram representing the difference between the parts of the wavefront used will be obtained in the overlap region. The asphericity noted in this interferogram will not be in name the same as that in the original, but will depend on it in a precise way. As an example, consider the effect of shearing two wavefronts containing identical first order coma given by $h = k \,.\, x \,.\, (x^2 + y^2)$, where h is the wavefront asphericity measured along the ray at the positional coordinate (x, y) (figure 3). Then the fringe pattern in the overlap region, given by the superposition of

$$h_r = k \,.\, (x+\alpha) \,.\, [(x+\alpha)^2 + y^2] \,,$$

and

$$h_e = k \,.\, (x-\alpha) \,.\, [(x-\alpha)^2 + y^2]$$

(where each wavefront is sheared a distance α along the x or shear axis), contains

$k \,.\, 4 \,.\, \alpha^2 \,.\, (x^2+y^2)$, change of focus;
$k \,.\, 2 \,.\, \alpha^2 \,.\, (x^2-y^2)$, astigmatism;
$k \,.\, 2 \,.\, \alpha^3$ change of path length.

In the same way it may be shown that pure spherical aberration shears into coma and a tilt, while pure astigmatism with axes along and perpendicular to the shear direction, produces in that direction a tilt which can be annulled by a lateral displacement of the apparatus. Astigmatism with axes skew to the shear direction produces, in addition, tilt perpendicular to the shear direction. Finally, when the wavefronts have revolution symmetry the interferogram represents an asphericity which is antisymmetrical about the centre O (figure 3).

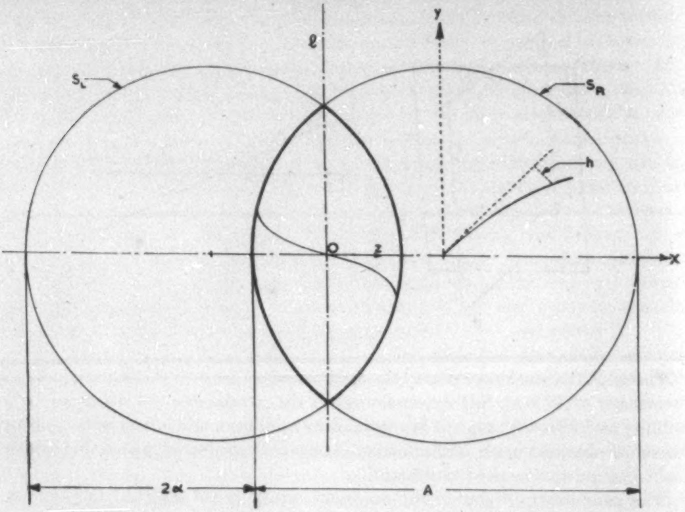


Figure 3.

The case of sheared paraboloids is of practical interest. The distance D between a paraboloid and a contacting sphere of the same polar radius of curvature ρ, measured along a diameter to that sphere at zonal radius z, is given by

$$D = \rho - \rho \,.\, \frac{[(1 - z^2/\rho^2)^{\frac{1}{2}} - (1 + z^2/\rho^2)^{\frac{1}{2}}]}{z^2/\rho^2}$$

$$= \frac{z^4}{8\rho^3} + \frac{7}{128} \cdot \frac{z^6}{\rho^5} + \dots \qquad \dots\dots(1)$$

Taking the leading fourth-power term only and shearing through plus and minus α as before, the interferogram retardation R is given by

$$R=\frac{z\alpha^3}{\rho^3}+\frac{z^3\alpha}{\rho^3}. \qquad \ldots\ldots(2)$$

Neglecting the term linear in z, the retardation across a meridian section is seen to be a cubic function of the position measured from the centre O (figure 3). By differentiating equation (2) with respect to α, it may be seen that the maximum amount of this cubic retardation occurs when the half-shear α is one-quarter of the linear aperture A. The maximum value of z will then be $A/4$, and the retardation at the interferogram edge is then $\frac{1}{256}\cdot\frac{A^4}{\rho^3}$. For a paraboloidal wavefront tested at its polar centre of curvature, this retardation in fringes will be $\frac{1}{2048}\cdot\frac{A}{F^3}\cdot\frac{1}{\lambda}$, where A is measured in the same units as the wavelength λ, and F is the focal ratio of the paraboloid. For an F.5 paraboloid of focal length 1 metre this gives about 1·6 fringes in a wavelength of 5000 A., and this retardation is of the same order as the fourth-power asphericity.

§ 5. CODA

Of the many methods of testing a wavefront, each has its particular advantages and limitations. The interferometer appears to have an advantage over ray methods in that it determines height differences directly rather than differences of slope, and a gain in simplicity over the usual methods of two-beam interferometry in that a separate comparison system is not required, Only experience will decide whether or not the advantages outweigh the concomitant limitations.

The technique is, however, applicable to any convergent wavefront. In its present form it would seem to have an astronomical application, in that it may be bolted to a telescope at the eyepoint to test the optical system under working conditions with a star as source. "Turned edge" is very easily seen in this way, as is also small magnitude rapid zonal error. Plate 2 shows the fringe pattern obtained under small shear from a nominally spherical speculum mirror which possessed a turned edge, with zonal error in addition to a scratched surface. Use may also be found for the instrument in the routine testing of an astronomical mirror during figuring operations.

§ 6. ACKNOWLEDGMENTS

The author wishes to express his thanks to Professor A. M. Tyndall, to Dr. C. R. Burch for unflagging interest and encouragement and to Mr. F. T. Bannister for assistance in the construction of the first instrument which excluded plasticene.

APPENDIX

Aberrations of skewed plane parallel plates

If a principal axis of a spherical wavefront, incident on a plane parallel plate of thickness t and refractive index μ, lies along the normal to the plate, then the longitudinal aberration S_ϕ of a ray at an angle ϕ to the plate normal is given by

$$S_\phi = \frac{t}{\mu}\left\{1-\left[\frac{1-\sin^2\phi}{1-\sin^2\phi/\mu^2}\right]^{\frac{1}{2}}\right\}.$$

For a plate of refractive index 1·5 the magnitude of this aberration per unit plate thickness is given in table 1.

Table 1

$\phi°$	S_ϕ/t
0	0
10	0·00568
20	0·02326
30	0·05429
40	0·10144
50	0·16824
60	0·25842
70	0·37413
80	0·51319
90	0·66667

In figure 4 the distance P_0P_ϕ is equal to S_ϕ where P_0 is the paraxial focus. Define now a new principal ray to be that ray at an angle ϕ to the plate normal, and let rays at $+\theta$ and $-\theta$ to it meet it in A_+ and A_- respectively.

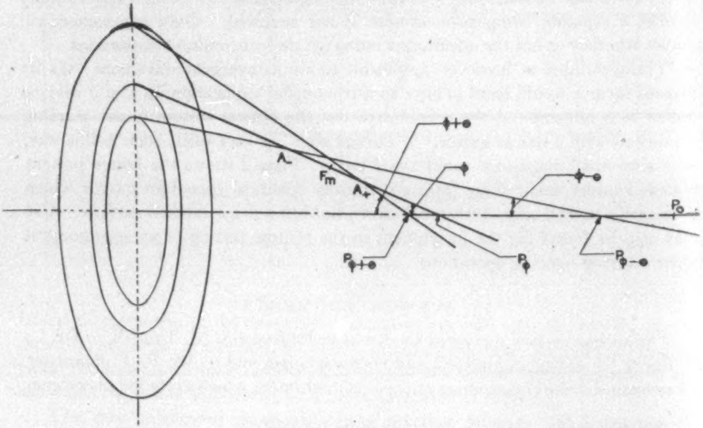


Figure 4.

If θ is reduced without limit the focus for rays paraxial with the new axis is determined at F_m, somewhere between A_+ and A_-. F_m is the location of the meridional focal line; the sagittal line, for fans out of the plane of the paper, is at P_ϕ. The distance F_mP_ϕ may be called the astigmatic interfocal distance.

With no restriction on θ the distance $\mathrm{A\_P}_{\phi}$ is given by

$$\mathrm{A\_P}_{\phi}=\sin(\phi-\theta)\,.\,\frac{[S_{\phi}-S_{\phi-\theta}]}{\sin\theta}$$

$$=[\sin\phi\,.\,\cot\theta-\cos\phi]\,.\left\{\theta\,.\,\frac{\partial S}{\partial\phi}-\frac{\theta^2}{\underline{|2}}\cdot\frac{\partial^2 S}{\partial\phi^2}+\ldots.+(-1)^{n-1}\,.\,\frac{\theta^n}{\underline{|n}}\cdot\frac{\partial^n S}{\partial\phi^n}+\ldots\right\}.$$

Expanding $\cot\theta$ in powers of θ this becomes

$$A\_P_{\phi}=\sin\phi\,.\,\frac{\partial S}{\partial\phi}$$

$$+\theta\,.\left\{-\cos\phi\,.\,\frac{\partial S}{\partial\phi}-\frac{1}{2}\,.\,\sin\phi\,.\,\frac{\partial^2 S}{\partial\phi^2}\right\}$$

$$+\theta^2\,.\left\{-\frac{1}{3}\,.\,\sin\phi\,.\,\frac{\partial S}{\partial\phi}+\frac{1}{2}\,.\,\cos\phi\,.\,\frac{\partial^2 S}{\partial\phi^2}+\frac{1}{6}\,.\,\sin\phi\,.\,\frac{\partial^3 S}{\partial\phi^3}\right\}$$

$$+\theta^3\,.\left\{\frac{1}{6}\,.\,\sin\phi\,.\,\frac{\partial S^2}{\partial\phi^2}+\frac{1}{6}\,.\,\cos\phi\,.\,\frac{\partial^3 S}{\partial\phi^3}-\frac{1}{24}\,.\,\sin\phi\,.\,\frac{\partial^4 S}{\partial\phi^4}\right\}$$

$$+\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;.\;. \qquad \ldots\ldots(3)$$

In this expression the term $\sin\phi\,.\,\partial S/\partial\phi$, which is independent of the aperture angle θ, gives the astigmatic interfocal distance $\mathrm{F}_m\mathrm{P}_{\phi}$.

These aberrations may be referred to the great sphere as follows for the purpose of error measurement in terms of fringes. The excess retardation of a ray at angle θ to the principal ray is given by

$$R_{\theta}-R_0=\int_0^{\theta}L(\theta)\,.\,\sin\theta\,.\,d\theta, \qquad \ldots\ldots(4)$$

where $\mathrm{L}(\theta)$ is the longitudinal aberration measured along the principal ray, and is the distance $\mathrm{A\_P}_{\phi}$. Writing equation (3) as

$$A\_P_{\phi}=A+\alpha\,.\,\theta+\beta\,.\,\theta^2+\gamma\,.\,\theta^3+\ldots.,$$

then equation (4) becomes

$$R_{\theta}-R_0=\theta^2\,.\,\frac{A}{2} \qquad \text{astigmatism;}$$

$$+\theta^3\,.\,\frac{\alpha}{3} \qquad \text{primary coma;}$$

$$+\theta^4\,.\left(\frac{\beta}{4}-\frac{A}{24}\right) \qquad \text{primary spherical aberration;}$$

$$+\theta^5\,.\left(\frac{\gamma}{5}-\frac{\alpha}{30}\right) \qquad \text{secondary coma;}$$

$$+\;.\;.\;.\;. \qquad \text{higher terms.} \qquad \ldots\ldots(5)$$

The values of these coefficients for a plate thickness unity, and refractive index 1·5, around $\phi=45°$ are given in table 2.

Table 2

$\phi°$	A	$\alpha/3$	$\beta/4-A/24$	$\gamma/5-\alpha/30$
33	0·136	−0·120	0·0563	−0·0002
35	0·155	−0·129	0·0565	+0·0015
37	0·176	−0·138	0·0562	0·0038
39	0·196	−0·147	0·0566	0·0067
41	0·219	−0·156	0·0554	0·0100
43	0·244	−0·165	0·0528	0·0139
45	0·270	−0·174	0·0504	0·0186
47	0·300	−0·181	0·0470	0·0241
49	0·327	−0·189	0·0429	0·0305
51	0·357	−0·197	0·0399	0·0375
53	0·390	−0·204	0·0302	0·0454
55	0·423	−0·211	0·0220	0·0543

From this table the aberrations introduced into a cone of semi-angle θ, incident at an angle ϕ to a plane parallel plate, may be determined. For example, from the first term of equation (5) it will be seen that the astigmatism introduced into a cone incident at $\phi°$ to a plate of thickness t is given as $\frac{A}{2} \, . \, \theta^2 \, . \, \frac{t}{\lambda}$ fringes, in a wavelength of λ. For an incidence of 45°, $A = 0{\cdot}27$ (table 2), and with an F.5 cone, plate thickness 0·5 cm., and wavelength 5000 Å., this astigmatism is about 13·5 fringes. Thus the compensating plates of the interferometer must be equal in thickness to the dividing plates to about 1/27 mm., in order to test such wavefronts to an accuracy of 1/10 fringe. With the same accuracy of thickness the second term of equation (5) shows that the associated primary coma is of the order of 1/100 fringe.

Consideration must also be given to errors in orientation of the interferometer plates. A change in plate orientation from 45° to 43° produces a change in the coefficient of A of 0·026 or a reduction of about 1·3 in the number of fringes of astigmatism. To test to an accuracy of 1/10 fringe will therefore require the plate orientations to be correct to about 9′ of arc.

The above tolerances are smaller than is actually necessary in the wavefront shearing interferometer, since the angular opening of the interference aperture is of necessity always less than that of the incident wavefront. Particularly is this so for large values of the shear, when also the coefficients of table 2 are larger.

From equation (3) it would appear that the higher order meridional aberrations do not increase excessively, but a rigorous proof of this has not been attempted. It will also be appreciated that the calculations consider only the meridian section of the wavefront.

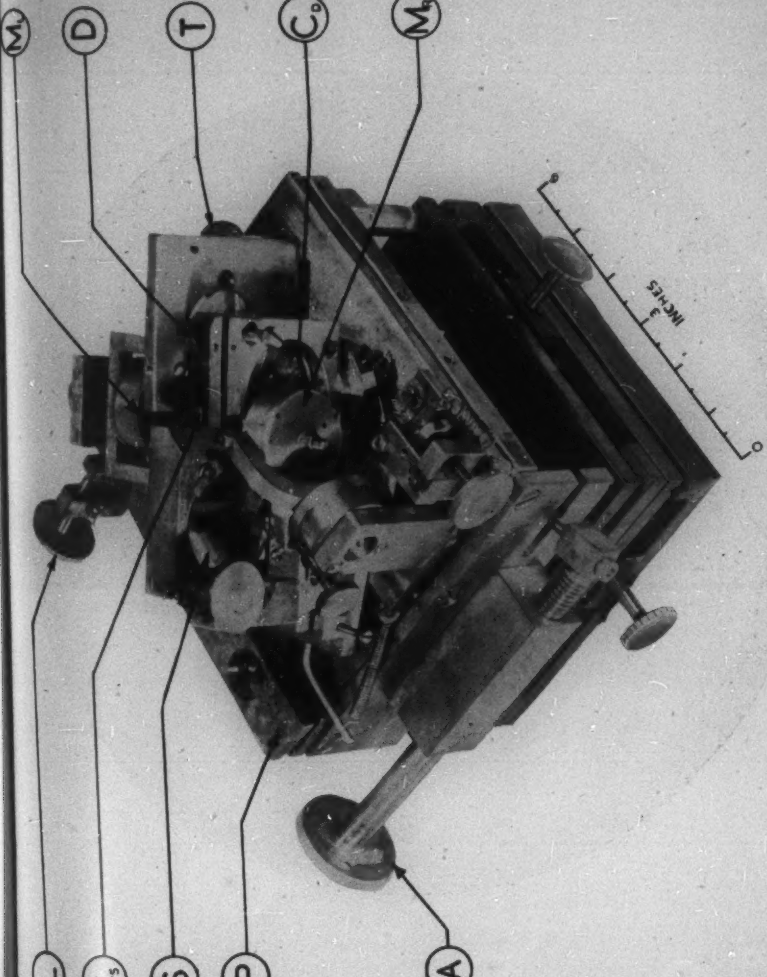
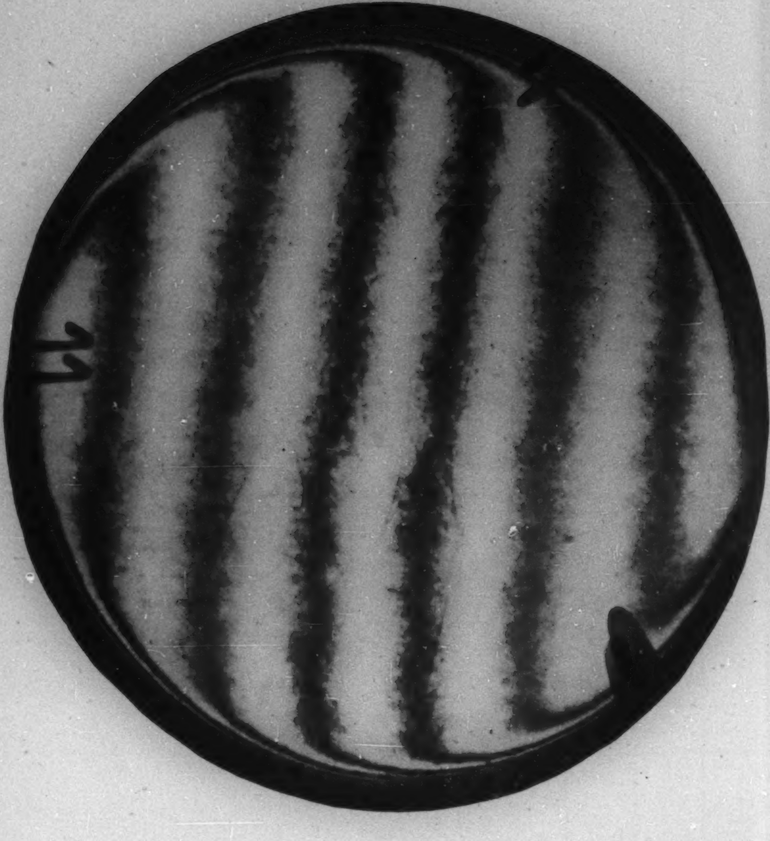

M_1
D
T
C_o
M_2
L
C_s
S
P
A
0
3
INCHES
6

Plate 1.



A TRANSPARENT-REPLICA TECHNIQUE FOR INTERFEROMETRY

By R. C. FAUST and S. TOLANSKY,
Royal Holloway College

MS. received 15 *May* 1947

ABSTRACT. A description is given of a transparent-replica technique which allows the surface of an opaque body to be examined interferometrically using the transmitted multiple-beam fringe pattern. This avoids the difficulties inherent in the technique of reflected fringes. A replica made from methyl methacrylate polymer is found to reproduce features both in extension and in depth to within close limits. Contours of only 40 A. in height are faithfully reproduced, and it is considered that a 10 A. change would be copied. The replica shows an overall shrinkage of the order of ½ per cent, but this is not a serious drawback.

The technique is tested on glass, mica and calcite, and is then applied to the examination of a coarsely polished metal surface, revealing features of interest. The new technique opens up further possibilities in the application of multiple-beam interference studies to the examination of the surfaces of polished metals. In particular, phase-effect errors are eliminated entirely.

§ 1. THE NEED FOR A REPLICA TECHNIQUE

RECENT developments in precision multiple-beam interferometry (Tolansky, 1946 a) for the study of surface topography are such that it has been found desirable to evolve a reliable replica technique in accordance with the following requirements.

The examination of an almost flat surface of a transparent body (e.g. a clear crystal, or a thin film) by multiple-beam interferometry is relatively simple, and fine sharp precision fringes are readily obtained (Tolansky, 1946 a) if the two surfaces concerned are suitably silvered and maintained at a separation no greater than a few wave-lengths of visible light. If a study is to be made of an appreciably concave surface, or of features lying below the general level of the specimen (for example deep etch pits), then the condition of close approach cannot be realized. Clearly, if a reliable negative replica cast can be prepared, the originally depressed features will now be elevated and can therefore be brought to within the requisite distance from the reference flat.

Of greater value is the application of a reliable replica technique to the examination of opaque surfaces, such as those of metals, as the following drawbacks are associated with the direct interferometric study of a metallic surface:—

(1) Because of opacity a back-reflection technique must of necessity be employed (Tolansky, 1945). If a low-power objective is in use, the working distance is sufficient to allow the insertion of a 45-degree reflector between the objective and the optical flat and no great difficulties are involved. This method is only of practical value for useful magnifications not exceeding × 50. If higher magnifications are required, the surface illumination practised in standard metallurgical microscopes must be resorted to. Here the reflector is between

the objective and the eyepiece, the objective performing the dual rôle of condensing light on to the specimen and at the same time forming an image of the illuminated surface. However, for multiple-beam interferometry the light falling on the specimen must be a strictly parallel beam at normal incidence. Hence the microscope objective must simultaneously produce a parallel beam from a small image formed at its rear focus and also act in its normal capacity as an image-forming lens. To carry out both functions correctly it would appear that a specially designed lens is necessary. The attainment of useful magnifications much in excess of ×50 is therefore associated with considerable experimental difficulties.

(2) In a forthcoming publication (Tolansky, 1947) it is shown that the visibility of the reflected fringe system is critically dependent upon the light-absorption of the silver film deposited upon the reference flat. If effective reflectivities as high as those permissible with transmission fringes are employed with the back-reflection technique, the fringe visibility is poor, and consequently the fringes are hard to detect. The visibility can be improved by lowering the reflectivity of the silver film on the flat, but this leads to an increased fringe width.

(3) The fact that the reflected fringes are fine dark lines on a broad bright background necessitates the use of monochromatic radiation. This is a drawback since, as has been shown before, considerable advantage in interpretation arises from the use of a mixed group of a small number of distinct wave-lengths. This restriction to monochromatic light means that a series of separate photographs must often be taken, each with a different wave-length, thus increasing both the labour and the difficulty of interpretation.

(4) Because of the above restriction the valuable "crossed-fringe" technique (Tolansky and Wilcock, 1946) cannot be adopted.

(5) A principal difficulty which can lead to serious major errors arises when metal surfaces, particularly alloys, are under examination. The point at issue is the question of differential phase change at reflection. Suppose the surface to possess a coarse, heterogeneous structure. Such a structure could arise for instance from any of the following: (*a*) different alloy constituents, (*b*) differential local ageing, (*c*) local polishing differences, (*d*) corrosion, (*e*) film formation. If the phase change at reflection varies over such local features then a serious error arises in interpretation since an optical change in phase can then be misinterpreted as a considerable metrical change in level. There would appear to be two ways of avoiding this error. One method would be to evaporate a fairly thick film of silver over the metal surface, and thus impose a uniform phase change (that for the silver) over the whole area. This appears attractive, but there may also be an error involved in so far as we are ignorant as to whether regional variations beneath the silver (e.g. corrosions) locally affect the phase change. (It is recalled that the silver film will be less than 1000 A. thick.) The second method for overcoming this difficulty is to develop a technique for making a reliable transparent replica casting, which eliminates phase effects.

Thus it is seen that if a transparent replica technique can be evolved, all the difficulties discussed will be removed and the simpler methods available

Figure 1.

Figure 2.

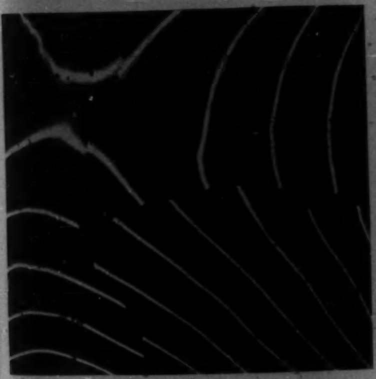


Figure 3.

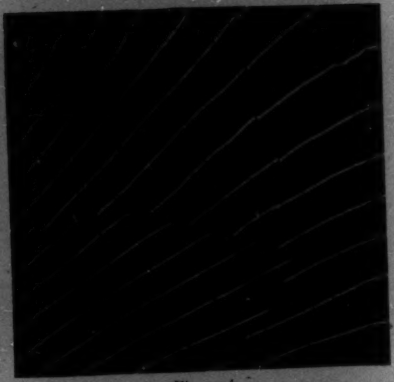


Figure 4.

Figure 5.

Figure 6.

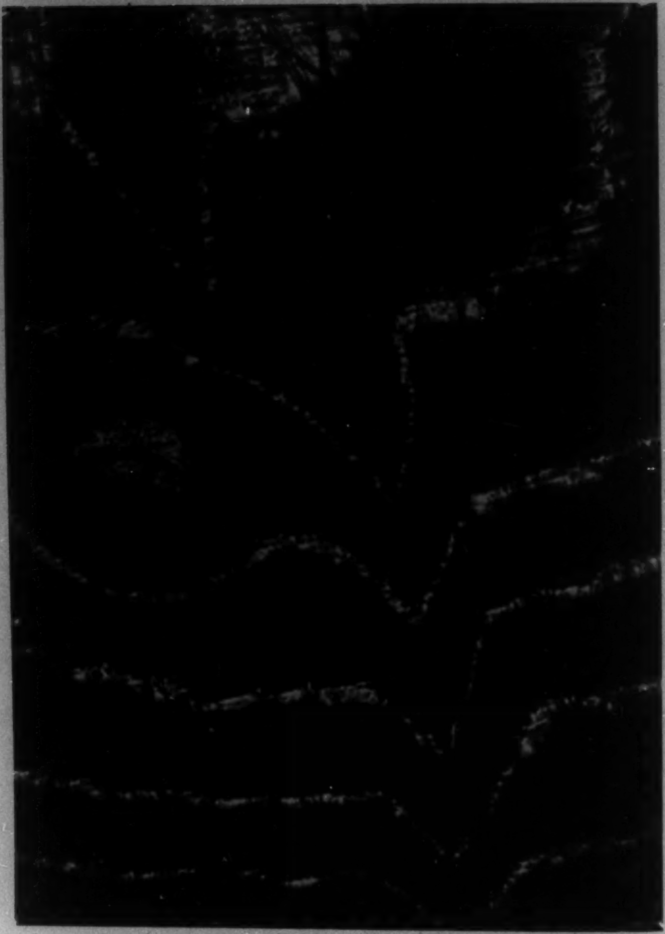


Figure 7.

for transparent bodies can be applied to the study of the surfaces of opaque solids. The realization of this possibility would be a considerable advantage.

Such a technique has been successfully developed, and this paper is largely devoted to a report on the reliability and faithfulness of reproduction realized by the procedure.

§ 2. THE REPLICA TECHNIQUE

A replica is required with the following characteristics:

1. It must be transparent.
2. Features in extension (i.e. across the surface), the dimensions of which are at least as small as 1/500 millimetre, must be reproduced with high fidelity.
3. Features in depth must be reproducible to within molecular dimensions.
4. As the replica is to be handled for adjustment it should be robust.
5. It should be capable of remaining in a vacuum without distorting.
6. A copious gas stream must not be liberated from the replica when in a vacuum, otherwise a silver coating of low light absorption cannot be deposited upon the surface.

Two materials which it was considered might meet these requirements were the I.C.I. plastic products Transpex I (unplasticized polymethyl methacrylate) and Transpex II (unplasticized polystyrene). Of these the methacrylate polymer is much to be preferred since it is far more readily degassed in the vacuum than the polystyrene, and all the measurements and photographs presented in this paper were made with Transpex I replicas. The material employed is in sheets about 4 mm. thick. From such a sheet a small piece is cut and washed with soap and warm water and then transferred to an oven where it is maintained at a temperature of about 170° c. At this temperature the plastic rapidly dries and is brought to a state suitable for moulding. Drying with a fabric is not attempted since frictional charges are readily built up, resulting in the collection of fibres and dust. The specimen, mounted in a suitable cement, is placed on an electric hot plate and brought to a temperature of 140° c. The plastic is placed on the specimen and then covered with a piece of hot plate glass. Upon the application of slight pressure the soft plastic flows and moulds itself on the one hand to the specimen surface and to the plate glass surface on the other. The plate glass is depressed until it finally rests on an accurately machined brass ring of constant height, this ring being in turn supported on a plane containing the specimen face. Consequently the finished replica approximates closely to a plane parallel sheet. This is desirable as the presence of an appreciable wedge-angle between the faces of the replica gives rise to ghost images and secondary fringes. The whole is allowed to cool slowly (at least an hour) to room temperature to avoid the setting up of internal strains. The replica and specimen are easily separated, the former then being coated with an evaporated silver film. Due to the flow characteristics of the plastic, the moulding pressure employed is quite small and it is not considered that specimen deformation is introduced.

The specimen is mounted, according to its character, in a suitable cement capable of withstanding the moulding temperature. For this purpose an artificial stone, Kaffir "D", has been found generally useful. An alternative investment is a phenol-formaldehyde thermosetting resin. Prior to the preparation of a replica it is helpful to clean the specimen by employing the well-known method of repeatedly stripping off collodion films.

§ 3. REPRODUCTION IN EXTENSION

To test the reliability of the reproduction in extension, casts were made from plane gratings ruled on metal. The replica gratings (which are quite robust and far superior to the usual Thorpe replicas found in teaching laboratories) were mounted on a spectroscope and the number of lines per inch determined with the sodium D lines in the usual manner. The values so obtained are compared below with the data marked on the original gratings.

	Lines per inch		
Original	2400	14200	17300
Replica	2412	14270	17370
Difference	+0·5%	+0·5%	+0·4%

Of particular interest is the fact that the 17,300-line metal grating showed symmetrical ghost lines, whose intensity increased with order number. Identical ghosts were observed with the replica grating. Furthermore, the definition and resolution of the replicas were in each case identical with those of the parent grating. It is clear that in these three cases the lateral structure of the specimen surface was very closely reproduced by the replica, the whole scale of the structure having undergone, however, a contraction of about $\frac{1}{2}$ per cent.

The $\frac{1}{2}$ per cent contraction arises from the difference between the thermal expansion of the metal and that of the plastic. The plastic tends to conform to the specimen structure down to the setting point (about 100° c.), at which state it will possess the structure of the thermally expanded metal at that temperature. As the coefficient of thermal expansion of the metal is about $0{\cdot}2 \times 10^{-4}$ per degree centigrade over the range 100° to 20° c., it contracts by about 0·2 per cent. The expansion coefficient of the plastic is higher, being about 1×10^{-4} per degree centigrade; consequently the plastic contraction is some 0·8 per cent. The difference between these two contractions, 0·6 per cent, represents the net shrinkage of the cooled plastic relative to the final cold state of the specimen. This figure is in close enough agreement with the measured values for the effective contraction. It follows that a sufficiently accurate correction can be made if 1×10^{-4} per degree centigrade be taken as a mean linear coefficient for the plastic. In the majority of cases this small correction can be disregarded.

§ 4. REPRODUCTION IN DEPTH

Whilst the many well-known replica techniques developed for electron microscopy indicate that faithful reproduction in extension might have been anticipated, such experience offers little evidence for the expected behaviour of a replica in terms of depth. It is this aspect which is the crucial one for interferometry, as it is in the determination of small heights and depths that the multiple-beam interference technique is so specifically powerful. It will now be shown that the reproduction in depth is very close indeed, and sufficient for many purposes. Replicas have been made of surfaces, the characteristics of which were already known from previous interference experiments, and the replicas then compared with the originals.

1. *Glass.* Figure 1 shows the multiple beam Fizeau fringes (λ 5461) formed when a replica of a piece of thin glass has been silvered and matched against a similarly silvered glass flat. The magnification is ×40. The fringes are typically those shown by glass. It has already been demonstrated elsewhere that the fine disrupted structure of the fringes is due to polish marks on the reference flat, this structure appearing only under critical illumination conditions.

2. *Mica.* It has previously been established that the steps appearing on the cleavage faces of muscovite are often small integral multiples of the 20-A. lattice spacing. A sample of muscovite was baked at 120° C. to drive off occluded water and a replica was then taken from a freshly cleaved face. Both the original and the copy were silvered and examined using Fizeau fringes. The characteristic cleavage steps were reproduced on the plastic, these steps being compared with the corresponding ones on the mica. It was known that a given cleavage step on mica is often constant in height to within a single molecular lattice over lengths of several millimetres. Measurements on the replicas showed that in this case also the step heights were true to within 5 A., the error in observation. The replica is so exact that it is quite easy to identify cleavage lines, and a precise comparison between the step heights on the mica and on the replica can be made. The following table is selected arbitrarily from such measurements, the step heights being in Ångström units. The experimental errors in each of the values quoted is of the order of 5 A.

Mica	385	505	1026	1280	1318	3237	4865
Replica	385	499	1020	1274	1302	3196	4822
Difference (%)		−1·2	−0·6	−0·5	−1·2	−1·3	−0·9

It is seen that over an extensive range the replica steps closely follow the original mica ones, but are consistently smaller than the latter by about 1 per cent. This shrinkage is not detected in the first step, being masked by the somewhat larger experimental errors. This 1 per cent shrinkage in depth is consistent with the shrinkage found in extension with the metal gratings, since the expansion coefficient of mica is considerably less than that of a metal.

Figures 3 and 4 show Fizeau fringes given by a sample of mica and by its replica. Figure 2 shows the same replica illuminated with the mercury yellow doublet instead of the green line (×50). As the fringe pattern is determined by the relative disposition between the silvered surfaces of the specimen and the reference flat, it is natural that the general contours from the mica and from the replica should differ. The important points for comparison are:

(*a*) The correspondence in the outlines of the cleavage steps.

(*b*) The heights of the corresponding steps, which are not influenced by flexure of the specimen. (The replica photographs are mirror images of the original mica, and raised features on the latter become depressions on the former.) The scratch marks in figure 2 are on the unsilvered face of the thin mica slip and are, therefore, not reproduced by the replica.

The fidelity of reproduction is very clearly shown not only by the numerical data but also by the clearly defined small step running in a nearly vertical direction in the right halves of figures 2 and 4. This step, which is only about $\lambda/40$ high, is nevertheless quite readily detectable. Evidence concerning the lower limit of reproducibility is provided by an interesting feature which has emerged from the examination of one particular cleavage line. Upon running along the length of this step on the mica a discontinuity in step height was found to occur, the height changing from 1320 to 1279 A. (Each of these values is perhaps in error by 3 A.) This change of 41 ± 6 A. corresponds to two molecular lattices. The striking fact is that at this point the replica exhibited a corresponding change of 34 ± 6 A. Thus notwithstanding the length of the long flexible polymer molecules, the flowing plastic contours the mica so critically that a change in height of only two mica lattices is almost exactly followed.

Figure 5 shows fringes given by another mica replica (× 50). The apparent black gaps between major "plane" areas arise from the occurrence of several close, narrow, descending steps, which result in a loss of light. The characteristic smooth continuuity of the mica fringes, formerly shown to indicate that mica cleaves true to a molecular plane over extensive areas, also appears on the replica. Indeed, despite experience and familiarity extending over several years in the examination of the fringes given by mica, we are quite unable to distinguish between fringes given by micas and those given by replicas, confusing both types completely if descriptive marks are obliterated.

These observations indicate that topographical features, even perhaps as small as 10 A. in height, will be reproduced with high fidelity.

3. *Calcite.* Figure 6 shows the fringes given by a replica taken from a cleavage face of a calcite crystal (× 50). This photograph is characteristic, closely simulating the calcite fringes already familiar to workers in this laboratory. It was not considered profitable to make measurements on calcite, since it has been established (Tolansky and Khamsavi, 1946) that the application of light pressure to a calcite cleavage surface leads to the gliding of crystal units, and this changes the height of certain cleavage steps. The reproduction is included to show that a mould can be taken of a relatively soft, friable surface.

4. *Polished steel.* Figure 7 is a reproduction of the fringes (× 70) given by a replica from a piece of steel which had been crudely polished and then buffed. The region selected includes a scratch mark about half a wave deep. It can be assumed from the preceding observations that the replica faithfully reproduces the metal surface topography, and this being the case, a number of points of interest emerge.

The considerable fringe width is evidence of surface irregularities, all the bright points in a given disrupted and broadened fringe being equidistant from the reference surface. The fringe broadening is greatest when the fringe direction is parallel to the clearly delineated polish scratches. The maximum fringe width in this particular photograph is about one fifth of an order, corresponding to polish marks having a height (depth) of the order of $\lambda/10$. It is obvious that sharp fringes can only be formed with highly polished surfaces.

A clearly distinguishable feature on this interferogram is the well-marked ridge. A ridge on the replica corresponds to a rut, or scratch on the original.

One can clearly discern that the metal is ploughed up a small fraction of a light wave on either side of the scratch mark, which is little more than 1 mm. long and perhaps 2000 A. deep. It is to be emphasized that these features recorded here are quite free from differential phase-change errors.

§ 5. CONCLUSION

The replica technique as described has been subjected to critical tests on transparent materials, whose topographical features are already well established, in order to obtain a measure of its reliability. These experiments demonstrate that the technique employed is suited to the purpose of examining the topography of opaque bodies, such as metals. The manifold requirements demanded for a faithful replica technique are met. Reproduction in extension is correct to well within the desired limits, but a general shrinkage of some ½ per cent in the whole contour takes place when a replica is taken from metal. For many studies such an effect is of little consequence. In depth reproduction it is surprisingly good. Features whose contours are smaller than 300 A. reproduce with an accuracy which is within the experimental error of 5 A., and a change in height of only 40 A. has been copied with fidelity. It is probable that a change of only 10 A. would be followed. The shrinkage effect takes place in depth as well as in extension and can be easily corrected for, if desired.

The methyl methacrylate used for the replica takes a high grade silver coating, exhibiting low absorption, from which it can be concluded that the final evolution of gas from the plastic surface must be small. The softness of the plastic permits it to be brought into intimate contact with a high grade optical flat, a somewhat risky procedure with the hard metal. This intimate contact leads to improved fringe definition, according to views already developed elsewhere.

Although the existence or otherwise of any creep effects in the replica has not been intensively studied, their influence, if any, must be slight, since observation has shown that a replica grating has remained unaltered over several months.

§ 6. ACKNOWLEDGMENT

This investigation was made possible with the aid of a maintenance grant from D.S.I.R. to R. C. Faust.

REFERENCES

Tolansky, 1945. *Proc. Roy. Soc.*, A, **184**, 41, 51.

Tolansky, 1946 a. *Proc. Roy. Soc.*, A, **186**, 261.

Tolansky, 1946 b. *Proc. Phys. Soc.*, **58**, 654.

Tolansky, 1947. *Proc. Zeeman Congress, Amsterdam,* 1946 (in the press).

Tolansky and Khamsavi, 1946. *Nature, Lond.*, **157**, 661.

Tolansky and Wilcock, 1945. *Nature, Lond.*, **157**, 583.

THE THEORY OF AN OSCILLATOR COUPLED TO A LONG FEEDER, WITH APPLICATIONS TO EXPERIMENTAL RESULTS FOR THE MAGNETRON

By C. DOMB,
Pembroke College, Cambridge
(sometime of Admiralty Signal Establishment)

MS. received 17 *December* 1946

§1. INTRODUCTION

The phenomena associated with an oscillator having coupled circuits in its output were dealt with many years ago in the early papers of van der Pol (1922) and others. It was there shown how under certain conditions two stable frequencies were possible, and the oscillator might start up on either. The purpose of the present paper is to deal with the case when the output consists of a length of feeder with a mismatch at the end. A good deal of attention has been focused on this problem recently because of the centimetre wave technique in which no R.F. amplification is possible, and feeders are often some hundreds of wavelengths long.

The problem falls naturally into two categories:

(*a*) When the feeder is sufficiently short for the time of forward and backward travel of the wave along it to be small compared with the time of rise of the oscillator; this may be termed the Static Case.

The feeder may be considered as an ordinary impedance element and the problem bears many similarities to the case of coupled circuits.

(*b*) When the feeder is sufficiently long for the propagation time to be of importance; this may be termed the Dynamic Case. The form of the initial frequency modulation before the oscillator settles down to its ultimate frequency is now of great interest, particularly in the case of pulsed transmissions. This is fundamentally different from anything in the classical theory.

The theory was developed for a negative resistance oscillator with a general characteristic. Experimental data were available for the magnetron, and an attempt was made to apply the theory using an empirical characteristic. The normal method of doing this would be to make detailed impedance measurements throughout the output circuit and hence determine the power output as a function of load. It is shown, however, how the static frequency pulling and power output curves for lengths of line varying in phase from 0 to 2π can be used to provide the data required.

§ 2. THE STATIC CASE

It is well known that many of the properties of oscillators can be accounted for by replacing the source of oscillations by a negative resistance. By this means we can deduce the frequency of the oscillations, but in order to determine the power output we must proceed to a further approximation. The condition for oscillation to be possible at a stable amplitude level is that the amount of power provided by the negative resistance must equal the power dissipated in the positive resistances in the circuit. If this source of power for any given amplitude level is greater than the power dissipated, the amplitude of the oscillations will increase. The simple assumption of a negative resistance is not sufficient to enable us to tell to what height the amplitude of the oscillations will rise. In order to do this we must assume a negative resistance characteristic—that is, a negative resistance which is a function of the voltage across it and which decreases when the voltage rises above a certain level. Whereas a simple negative resistance corresponds to a relation between current and voltage of the form $i=-av$, what is required is a relation $i=\chi(v)$ where $\chi(v)$ is of the form i_0-av for small v but flattens out as v increases. An example of this type is shown in figure 1. There will be one point on the curve at which the power provided by the source during a cycle of oscillation is exactly balanced by the power dissipated, and this is the height to which the amplitude of the oscillations will rise.

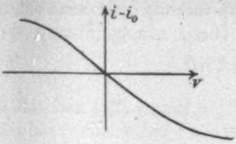


Figure 1.

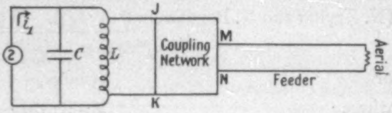


Figure 2.

Frequency and amplitude of a steady solution

The system with which we shall be concerned is shown diagrammatically in figure 2. However, it is not difficult to deduce the amplitude and frequency of steady sinusoidal oscillations for any general linear output network. Let v be the voltage across the generator. Then the differential equations for any network can be put in the form

$$\left.\begin{aligned} v &= z_{11}i_1 + z_{12}i_2 + \dots + z_{1n}i_n \\ 0 &= z_{21}i_1 + z_{22}i_2 + \dots + z_{2n}i_n \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0 &= z_{n1}i_1 + \dots \qquad + z_{nn}i_n \end{aligned}\right\} \qquad \dots\dots(1)$$

where z_{ij} is a function of the operator $p\equiv d/dt$.

Solving for i_1 in terms of v, we obtain

$$p^m\Delta(p)i_1 = p^m\Delta_1(p)v, \qquad \dots\dots(2)$$

where

$$\Delta(p) = \begin{vmatrix} z_{11} \dots z_{1n} \\ z_{21} \dots z_{2n} \\ \cdot\ \ \cdot\ \ \cdot\ \ \cdot \\ z_{n1} \dots z_{nn} \end{vmatrix} \qquad \Delta_1(p) = \frac{\partial\Delta}{\partial z_{11}} = \begin{vmatrix} z_{22} \dots z_{2n} \\ \cdot\ \ \cdot\ \ \cdot\ \ \cdot \\ z_{n2} \dots z_{nn} \end{vmatrix}$$

and the factor p^m is inserted to eliminate any inverse power of p.

Also, as mentioned above, we assume a relation

$$i_1 = -\chi(v)\text{*}. \qquad \ldots\ldots(3)$$

Therefore

$$p^m \Delta(p)\chi(v) + p^m \Delta_1(p) v = 0. \qquad \ldots\ldots(4)$$

Equation (4) is a non-linear equation many examples of which have been considered by Appleton (1923), van der Pol (1934) and others. We shall be concerned only with cases in which a solution exists which is approximately sinusoidal. Assume a solution

$$v = A\cos\omega t = \operatorname{Re}(Ae^{i\omega t}). \qquad \ldots\ldots(5)$$

To deal with the non-linear terms of (4), i.e. those involved in $\chi(v)$, we follow a procedure similar to that of van der Pol. We assume that $\chi(A\cos\omega t)$ can be expanded as a Fourier Series,

$$\chi(A\cos\omega t) = a_0 + a_1\cos\omega t + \ldots.\, a_n \cos n\omega t + \ldots. \qquad \ldots\ldots(6)$$

We now ignore all harmonic terms assuming that they can be neglected in comparison with the fundamental term. The detailed justification of this step is complicated, and has been considered at some length recently by Russian authors (N. Kryloff and N. Bogoliouboff (1943)). We then obtain on substituting in (4)

$$(i\omega)^m \Delta(i\omega) a_1 + (i\omega)^m \Delta_1(i\omega) A = 0 \qquad \ldots\ldots(7)$$

or

$$Z(i\omega) = -A/a_1, \qquad \ldots\ldots(8)$$

where

$$Z(i\omega) \equiv \Delta(i\omega)/\Delta_1(i\omega)$$

is the output impedance presented to the generator.

If we equate real and imaginary parts of (8) we obtain

$$\operatorname{Im} Z(i\omega) = 0 \qquad \ldots\ldots\ (9)$$

$$\operatorname{Re} Z(i\omega) = R = -A/a_1. \qquad \ldots\ldots(10)$$

Equation (9) determines the possible frequencies.

Equation (10) can be shown to be an expression of the conservation of energy. For by (6)

$$\overline{\cos\omega t\,\chi(A\cos\omega t)} = \tfrac{1}{2}a_1,$$

so that (10) may be written in the form

$$-\overline{v\chi(v)} = -\overline{A\cos\omega t\,\chi(A\cos\omega t)} = A^2/2R = \overline{v^2}/R. \qquad \ldots\ldots(11)$$

The left-hand side of (11) is the mean power supplied by the source during a cycle of oscillation, and the right-hand side is the mean power dissipated in the network. It will be noted that to this order of approximation the frequency is independent of the form of the characteristic of the oscillator.

Initial rise of the oscillations

So far we have only been concerned with the resultant amplitude of the steady oscillation. But the procedure of van der Pol can also be employed, in the case of a sinusoidal oscillation, to deduce a differential equation for the initial curve of rise of the oscillations. Instead of the substitution (5) we now use

$$v = V(t)\cos\omega t = \operatorname{Re}[v(t)e^{i\omega t}] \qquad \ldots\ldots(12)$$

---

[*] The minus sign occurs since the current is out from the generator and not in towards it.

and assume that dV/dt can be neglected in comparison with ω, and similarly for higher derivatives. This method will be of great importance in §3. We now apply it, by way of an illustrative example, to the simple circuit of figure 3.

It is easy to show that in this case

$$\frac{d^2v}{dt^2}+\left[\frac{1}{CR}+\frac{1}{C}\chi'(v)\right]\frac{dv}{dt}+\omega_0^2v=0 \qquad \ldots\ldots(13)$$

$$\omega_0^2=1/LC.$$

Using the above procedure, and neglecting harmonics as above, we obtain

$$\frac{dV}{dt}+\frac{1}{C}\left[\frac{V}{2R}-\frac{g(V)}{V}\right]=0, \qquad \ldots\ldots(14)$$

where $g(V)=-\overline{v\chi(v)}=-\overline{V\cos\omega t\,\chi(V\cos\omega t)}$ = power output. The simplest possible non-linear form for $\chi(v)$ is that used by van der Pol, $\chi(v)=i_0-b_1v+b_3v^3$;

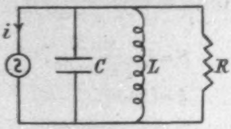


Figure 3.

in this case $g(V)$ equals $b_1V^2/2-3b_2V^4/8$. We quote the complete solution in this case from van der Pol's paper (1934).

$$V=\sqrt{\frac{4b_1'}{b_3'[1+e^{-b_1'(t-t_0)}]}}, \qquad b_1'=\frac{b_1}{C}-\frac{1}{CR}, \qquad \ldots\ldots(15)$$

$$b_3'=3b_3/C.$$

Output circuits

In nearly all practical transmitters, the tuned circuit of the oscillator is coupled to the aerial, which radiates the power, by means of a reactive network. In the case of triode oscillators, the coupling is usually purely electromagnetic; in the case of a magnetron it usually consists of the coupling loop and the waveguide output system.

For any reactive network the solution of (1) at any given frequency ω, can be put in the form

$$Z=z_{11}+\sum_{i=2}^{n}\frac{z_{1i}\Delta_i(i\omega)}{\Delta_1(i\omega)}, \quad \text{where} \quad \Delta_i=\partial\Delta/\partial z_{1i}.$$

All the z_{ij} are pure reactances except $z_{nn}=z$ which is the impedance presented by the aerial feeders. This may be reduced to

$$Z=z_{11}+A(1+iBz)/(z+iC), \qquad \ldots\ldots(16)$$

where A, B and C are purely real. An exactly analogous procedure is possible for admittances.

As an illustrative example in the case of purely electromagnetic coupling, if M is the mutual inductance and N the self-inductance of the coupling loop

$$Z=z_{11}+\omega^2M^2/(z+i\omega N), \qquad \ldots\ldots(17)$$

so that $A=\omega^2M^2$, $B=0$, $C=\omega N$.

In general, A, B and C are functions of ω, but in most practical cases, such as the one just quoted, they are slowly varying functions of ω, so that if ω varies by only a small fraction of itself they may be regarded as constants.

We now consider in detail the circuit of figure 2.

As our equivalent circuit has been taken with elements in parallel it is convenient to use admittances. We assume that the impedance at the aerial is a resistance $r=1/g$. Let Y_0 be the characteristic admittance of the feeder and l its length. Then admittance across MN presented by the feeder is given by

$$y=Y_0\frac{g\cos\theta+iY_0\sin\theta}{Y_0\cos\theta+ig\sin\theta}=Y_0\frac{\alpha\cos\theta+i\sin\theta}{\cos\theta+i\alpha\sin\theta}, \qquad \ldots\ldots(18)$$

where $\theta=\omega l/c=2\pi l/\lambda,\quad \alpha=g/Y_0=$ standing wave ratio.

Taking real and imaginary parts and writing $y=G+iS$ we find

$$G\equiv\frac{1}{R}=\frac{2\alpha Y_0}{1-\alpha^2}\frac{1}{\xi+\cos 2\theta}, \qquad \ldots\ldots(19)$$

$$S=Y_0\frac{\sin 2\theta}{\xi+\cos 2\theta}, \qquad \ldots\ldots(20)$$

where $\xi=(1+\alpha^2)(1-\alpha^2)$.

If we ignore the coupling network completely (i.e. connect straight to the feeder) the frequency would satisfy the relation

$$\omega C-1/\omega L+S=0$$

using (9), or

$$\Delta\omega=-\frac{Y_0}{2C}\frac{\sin 2\theta}{\xi+\cos 2\theta}=\frac{Y_0}{2C}\frac{\sin 2\phi}{\xi-\cos 2\phi}, \qquad \ldots\ldots(21)$$

where

$$\Delta\omega=\omega-\omega_0,\quad \phi=\theta+\pi/2.$$

The power output would be given by equation (11) with

$$\frac{1}{R}=\frac{2\alpha Y_0}{1-\alpha^2}\frac{1}{\xi-\cos 2\phi}. \qquad \ldots\ldots(22)$$

It will be shown a little later that so long as the reactances in the coupling network can be considered constant over the small band of frequencies covered, which is nearly always the case in practice, the coupling network produces no change in the shape of the curves given by (21) and (22), but merelyshifts them or magnifies them. Hence the functions

$$\sin 2\phi/(\xi-\cos 2\phi) \qquad \ldots\ldots(23)$$

and

$$1/(\xi-\cos 2\phi) \qquad \ldots\ldots(24)$$

may be thought of as basic functions in connection with the frequency and power output. They are drawn as functions of ϕ, for various values of ξ in figures 4 and 5.

We also enumerate for convenience several properties of the curves (23) and (24). The maximum of (24) occurs when $\phi=0$ and then $(23)=0$. Similarly the minimum of (24) occurs when $\phi=\pi/2$ and again $(23)=0$. The maximum and minimum of (23) occur when $\cos 2\phi=1/\xi$, and have values

$$\pm\frac{1}{\sqrt{\xi^2-1}}=\pm\frac{1-\alpha^2}{2\alpha}. \qquad \ldots\ldots(25)$$

The slope at the origin of (23) is given by

$$m_0 = \frac{2}{|\xi| - 1}. \qquad \ldots\ldots(26)$$

We now deal with a general reactive coupling network and prove the statement made previously. As was shown above, the admittance across JK facing toward the aerial can be put in the form

$$Y = A\,\frac{1 + i\beta y/Y_0}{y + i\gamma Y_0} \qquad \ldots\ldots(27)$$

where A, β, γ are real and assumed constant. Substituting the value of Y from (18) and writing $Y = G + iS$ we obtain after a little simplification

$$G = \frac{A}{Y_0}\,\frac{2\alpha}{1-\alpha^2}\,\frac{1+\beta\gamma}{(1+\gamma^2)\xi + 2\gamma\sin 2\theta - (1-\gamma^2)\cos 2\theta} \qquad \ldots\ldots(28)$$

and

$$S = \frac{A}{Y_0}\,\frac{(\beta-\gamma)\xi - (\beta+\gamma)\cos 2\theta - (1-\beta\gamma)\sin 2\theta}{(1+\gamma^2)\xi + 2\gamma\sin 2\theta - (1-\gamma^2)\cos 2\theta}. \qquad \ldots\ldots(29)$$

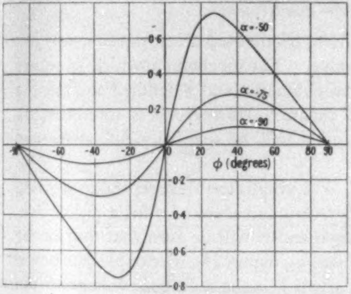


Figure 4.

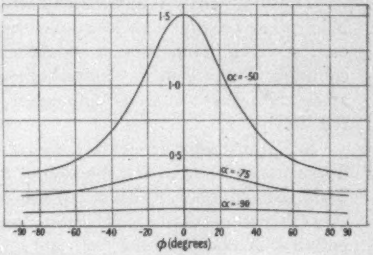


Figure 5.

Write $\gamma = \tan\lambda$; then

$$(1+\gamma^2)\xi + 2\gamma\sin 2\theta - (1-\gamma^2)\cos 2\theta = (1+\gamma^2)[\xi - \cos 2(\theta+\lambda)],$$

and it is now easy to show that

$$G = \frac{A}{Y_0}\,\frac{1+\beta\gamma}{1+\gamma^2}\,\frac{2\alpha}{1-\alpha^2}\,\frac{1}{\xi - \cos 2(\theta+\lambda)} \qquad \ldots\ldots(30)$$

and

$$S = \frac{A}{Y_0}\left[\frac{\beta-\gamma}{1+\gamma^2} - \frac{1+\beta\gamma}{1+\gamma^2}\,\frac{\sin 2(\theta+\lambda)}{\xi - \cos 2(\theta+\lambda)}\right]. \qquad \ldots\ldots(31)$$

The frequency and power output are given by equations similar to (21) and (22).

If we plot the right-hand side and left-hand side of (21) as functions of ω, the intersections of the resulting curves determine the possible frequencies. It is more convenient to plot them as functions of θ. Then the right-hand side is independent of the length of guide, and the left-hand side $= c(\theta - \theta_0)/l$ is a straight line whose slope is inversely proportional to l. For sufficiently small l the curves intersect in only one point. When l increases beyond the critical value given by

$$c/l = Y_0 m_0 / 2C \qquad \ldots\ldots(32)$$

three frequencies become possible for certain values of the angle θ_0, or for certain phase lengths of the feeder. As l increases still further five frequencies become possible, and so on; this question is discussed in greater detail in a companion experimental paper by B. W. Lythall (1946).

Analysis of experimental data

If the expressions (30) and (31) are examined closely, it will be seen that a considerable similarity exists between the coefficients of the basic frequency and power functions. In fact, if we are given the curve of S as a function of θ, and α is known, we can deduce the value of G at any point of the curve. In practice we do not know S but $\Delta\omega = -S/2C$; hence we can deduce $G/C = 1/CR$ at any point of the curve. Thus using (25) we deduce from (31) that

$$p = (\Delta\omega)_{\max} - (\Delta\omega)_{\min} = \frac{1-\alpha^2}{2C}\frac{A}{Y_0}\frac{1+\beta\gamma}{1+\gamma^2} \qquad \ldots\ldots(33)$$

and hence from (30)

$$\frac{1}{CR} = p\left(\frac{2\alpha}{1-\alpha^2}\right)^2 \frac{1}{\xi - \cos 2(\theta+\lambda)}. \qquad \ldots\ldots(34)$$

The above analysis applies satisfactorily to triode oscillators, the function $\chi(v)$ being related to the characteristics of the valve. The mechanism which produces oscillations in the magnetron is different in character, and a detailed analysis is a complicated problem in electromagnetic theory. However it is interesting to apply the present analysis empirically, and it will be seen to yield satisfactory results.

In figure 6 the experimental curves of B. W. Lythall (1946) are reproduced giving frequency and power as a function of θ for various values of α for a magnetron. Theoretical frequency curves are also drawn and it will be seen that they fit the experimental ones quite well. From these curves, using (34), we can plot the power output P as a function of $1/CR$; and similarly if V is the R.F. peak voltage we can plot $V\sqrt{C}$ as a function of $1/CR$. This is done in figure 7 for the given data. Hence we can plot P as a function of $V\sqrt{C}$, and determine, except for a constant, the function $g(V)$ of equation (14). This is done in figure 8, and can be used in the problem of the dynamic case.

The functions in figure 7 and figure 8 can be fitted over a good deal of their range by a very simple empirical formula; this has been previously referred to by American authors. It can be most easily demonstrated by plotting V, the R.F. peak voltage, against I, the R.F. peak current. Actually we plot $\sqrt{2PCR} = V\sqrt{C}$ against $\sqrt{2P/CR} = I/\sqrt{C}$ in figure 9, and will it be seen that the resulting points approximate quite closely, over a good deal of their range, to a straight line. Hence we may write

$$V = V_0 - R_0 I \qquad \ldots\ldots(35)$$

where, from figure 9,

$$\omega C R_0 = 44 \quad V_0\sqrt{C} = 8 \text{ ($\Delta\omega$ being measured in Mc./sec.)} \qquad \ldots\ldots(36)$$

and

$$g(V) = (V/2R_0)(V_0 - V) \qquad \ldots\ldots(37)$$

over the region where (35) is satisfied. This formula should not be extrapolated beyond the region where it is checked by experiment.

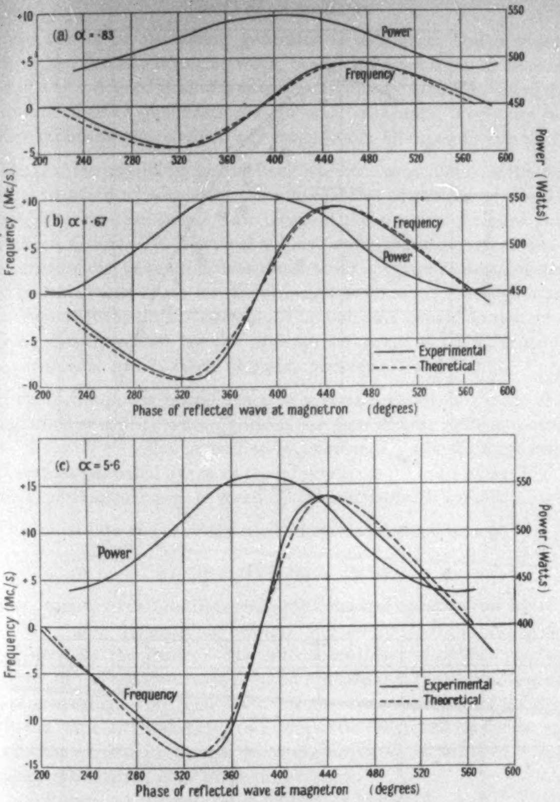


Figure 6.

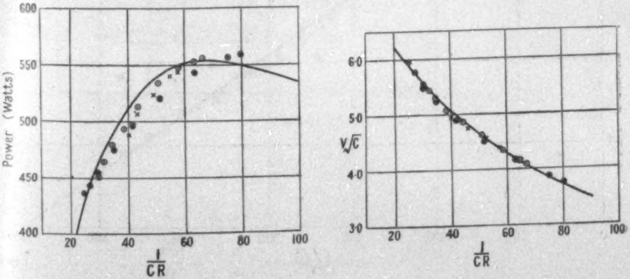


Figure 7.

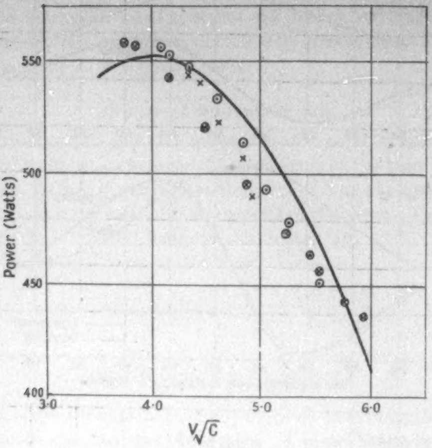


Figure 8.

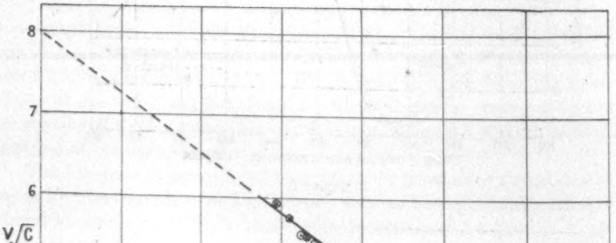


Figure 9.

The Q of an oscillator

The problem of an oscillation generator is non-linear, and is much more complicated than that of a linear network, to which we can apply an unambiguous definition of a Q. A good deal of confusion can thus be caused by using the term loosely without further qualification as to what is meant. There are certainly three different phenomena to which a Q is applicable, all depending on the load.

(1) The Q of the oscillating system itself without the oscillator; if the oscillator were switched off suddenly, the rate of decay of the power would enable this Q to be determined. In terms of the simple equivalent circuit of figure 3, $Q=1/\omega CR$. Typical values for the experimental results on the magnetron can be read off from the abscissa of figure 7 (assuming the guide to be sufficiently short for the propagation time to be negligible in comparison with the time of decay). The units are chosen so that we obtain Q by dividing 3000 by the appropriate value of $1/CR$; hence for the experimental points taken Q varies between 37 and 120.

(2) If an oscillation is taking place at a certain stable amplitude A and is disturbed at time t_0 to amplitude $A+x$ then at time t the amplitude will be $A+xe^{-\delta(t-t_0)}$; this gives rise to another Q. For the magnetron, this Q may be estimated by use of the empirical results of equation (35). Over the experimental range covered it varies between 20 and 32.

(3) The initial rate of rise of the oscillations also determines a Q.

§ 3. THE DYNAMIC CASE

We now propose to deduce rigorously the complete equations for the circuit of figure 2. We ignore the coupling network and assume that the output circuit is directly coupled to the feeder; this introduces some simplification, and the considerations of the the static case seem to indicate that the assumption involves little loss of generality. We then show that the steady solutions of the equations are exactly the same as those deduced in the static case; and we derive approximations which are useful in the case when the amplitude modulation is small and the frequency modulation is of prime importance.

Let the voltages and currents be as labelled in figure 10.

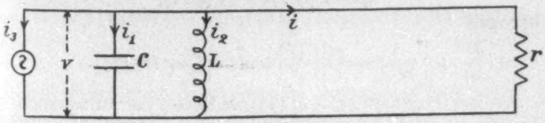


Figure 10.

Then
$$v=L\frac{di_2}{dt}=\frac{1}{C}\int i_1 dt$$

and
$$i_3=-(i+i_1+i_2)=\chi(v).$$

Hence
$$\frac{d^2v}{dt^2}+\frac{1}{C}\frac{d}{dt}[\chi(v)]+\omega_0{}^2v=-\frac{1}{C}\frac{di}{dt}. \qquad \text{......(38)}$$

From transmission line theory we have

$$\left.\begin{aligned} v &= f(t) - \rho f(t-\tau), \\ Z_0 i &= f(t) + \rho f(t-\tau), \end{aligned}\right\} \qquad \ldots\ldots(39)$$

where $\alpha = \dfrac{Z_0}{r} = \dfrac{g}{Y_0}$ as above, $\rho = \dfrac{1-\alpha}{1+\alpha}$,

τ = propagation time forwards and backwards along the line. We deal with equation (38) by the method described in §2.

Write

$$\left.\begin{aligned} v &= \mathrm{Re}\,(Ve^{i\omega_0 t}) = v_1 \cos\omega_0 t - v_2 \sin\omega_0 t \\ &= |V| \cos(\omega_0 t + \eta), \\ i &= \mathrm{Re}\,(Ie^{i\omega_0 t}), \\ f(t) &= \mathrm{Re}\,[F(t) e^{i\omega_0 t}], \end{aligned}\right\} \qquad \ldots\ldots(40)$$

where $|V|$ and η are slowly varying in comparison with ω_0.

We expand $\chi(v) = \chi[|V| \cos(\omega_0 t + \eta)]$

in terms of cosine harmonics of $(\omega_0 t + \eta)$ and retain only the first harmonic. Thus we write $\chi(v)$ as approximately equal to

$$-\frac{2g(|V|)}{|V|} \cos(\omega_0 t + \eta),$$

where $g(|V|)$ is defined by (14). Hence

$$\frac{d}{dt}[\chi(v)] = -2\frac{d}{dt}\left[\frac{g(|V|)}{|V|}\right] \cos(\omega_0 t + \eta) + (\omega_0 + \eta') \frac{2g(|V|)}{|V|} \sin(\omega_0 t + \eta),$$

and neglecting terms of order $\dfrac{\eta'}{\omega_0}$, $\dfrac{1}{\omega_0}\dfrac{d|V|}{dt}$ we may write this as

$$\mathrm{Re}\,\frac{2i\omega_0 g(|V|)}{|V|^2} Ve^{i\omega_0 t}.$$

Proceeding similarly with the other terms of (38) we obtain, after a little simplification,

$$\frac{dV}{dt} - \frac{g(|V|)}{C|V|^2} V = -\frac{1}{2C} I. \qquad \ldots\ldots(41)$$

Applying the substitution (40) to (39), we have

$$\left.\begin{aligned} V &= F(t) - \rho e^{-i\omega_0 \tau} F(t-\tau), \\ Z_0 I &= F(t) + \rho e^{-i\omega_0 \tau} F(t-\tau). \end{aligned}\right\} \qquad \ldots\ldots(42)$$

Hence $Z_0 I = F(t) + \rho e^{-i\omega_0 \tau} F(t-\tau)$,

and (41) becomes

$$\frac{dV}{dt} + \left[\frac{1}{2CZ_0} - \frac{g(|V|)}{C|V|^2}\right] V = -\frac{\rho}{CZ_0} e^{-i\omega_0 \tau} F(t-\tau). \qquad \ldots\ldots(43)$$

If we take real and imaginary parts of (43) we have a pair of equations which in conjunction with the first equation of (42) completely determine the form of the oscillation generated. Actually although this form of the equation is convenient for integration on a differential analyser, it is difficult to see the physical significance of the various terms. For the purpose of physical approximation it is more useful to transform the equation into a pair involving amplitude and phase explicitly. Taking the conjugate complex of (43),

$$\frac{d\bar{V}}{dt} + \left[\frac{1}{2CZ_0} - \frac{g(|V|)}{C|V|^2}\right] \bar{V} = -\frac{\rho}{CZ_0} e^{i\omega_0 \tau} \bar{F}(t-\tau). \qquad \ldots\ldots(44)$$

Multiplying (43) by $\overline{V}$, (44) by V and adding,

$$\frac{d}{dt}(|V|^2)+2(|V|^2)\left[\frac{1}{2CZ_0}-\frac{g(|V|)}{C|V|^2}\right]=-\frac{2\rho}{CZ_0}[v_1k_1+v_2k_2].$$

Similarly subtracting,

$$2i\frac{d}{dt}\tan^{-1}\frac{v_2}{v_1}=-\frac{2i\rho}{CZ_0}\frac{v_1k_2-v_2k_1}{|V|^2},$$

where

$$e^{-i\omega_0\tau}F(t-\tau)=k_1+ik_2=K(t). \qquad \ldots\ldots(45)$$

These equations can be written as

$$\left.\begin{aligned}\frac{d|V|}{dt}+\left[\frac{|V|}{2CZ_0}-\frac{g(|V|)}{C|V|}\right]&=-\frac{\rho}{CZ_0}[k_1\cos\eta+k_2\sin\eta],\\ \frac{d\eta}{dt}&=-\frac{\rho}{CZ_0|V|}[k_2\cos\eta-k_1\sin\eta].\end{aligned}\right\} \qquad \ldots\ldots(46)$$

As a check on the correctness of equations (46) we find the values of amplitude and frequency for which steady solutions are possible, and compare them with the results obtained in the discussion of the static case. The conditions for a steady solution are

$$d\eta/dt=\Delta\omega, \quad d|V|/dt=0.$$

Hence from (42)

$$F=\frac{V}{1-\rho e^{-i\omega\tau}}, \quad \omega=\omega_0+\Delta\omega.$$

Substituting in (46),

$$\Delta\omega=-\frac{\rho}{CZ_0}\operatorname{Im}\left(\frac{e^{-i\omega\tau}}{1-\rho e^{-i\omega\tau}}\right)=-\frac{1}{2CZ_0}\frac{\sin\omega\tau}{\xi+\cos\omega\tau}, \qquad \ldots\ldots(47)$$

where

$$\xi=-(1+\rho^2)/2\rho=(1+\alpha^2)/(1-\alpha^2)$$

and

$$\frac{|V|}{2CZ_0}-\frac{g(|V|)}{C(|V|)}=-\frac{\rho|V|}{CZ_0}\operatorname{Re}\left(\frac{e^{-i\omega\tau}}{1-\rho e^{-i\omega\tau}}\right)$$

or

$$\frac{g(|V|)}{|V|^2}=\frac{1}{2Z_0}\frac{2\alpha}{1-\alpha^2}\frac{1}{\xi+\cos\omega\tau}. \qquad \ldots\ldots(48)$$

Equations (47) and (48) will be found to be analogous to (21) and (22).

Approximations to the equations

Equations (46) are cumbersome, even for numerical integration, because of the complex difference relations. From the physical point of view, the best method of solution would appear to be one of successive approximations. Often only a qualitative picture is required for any given conditions, and then the first approximation is usually sufficient. Also in practice, the amplitude modulation is often small, so that only the frequency modulation need be considered to the first order.

From the first equation of (42) we deduce that

$$F(t)=V(t)+\rho e^{-i\omega_0\tau}V(t-\tau)+\rho^2e^{-2i\omega_0\tau}V(t-2\tau)+\ldots.$$

the series continuing until $V(t-n\tau)=0$.

Hence

$$\rho K(t)=\rho e^{-i\omega_0\tau}V(t-\tau)+\rho^2e^{-2i\omega_0\tau}V(t-2\tau)+\ldots.$$

and

$$\frac{\rho K(t)}{V}=\rho e^{-i[\omega_0\tau+\eta(t)-\eta(t-\tau)]}\left|\frac{V(t-\tau)}{V(t)}\right|+\rho^2e^{\;i[2\omega_0\tau+\eta(t)-\eta(t-2\tau)]}\left|\frac{V(t-2\tau)}{V(t)}\right|$$
$$+\ldots.\quad+\rho^re^{-i[r\omega_0\tau+\eta(t)-\eta(t-r\tau)]}\left|\frac{V(t-r\tau)}{V(t)}\right|+\ldots. \qquad \ldots\ldots(49)$$

We now approximate by assuming in (49) that all terms such as $\left|\frac{V(t-r\tau)}{V(t)}\right|$ can be put equal to unity. We put the resulting function equal to $H(t)$, so that

$$H(t)=h_1(t)+ih_2(t)=\rho e^{-i[\omega_0\tau+\eta(t)-\eta(t-\tau)]}+\rho^2e^{-i[2\omega_0\tau+\eta(t)-\eta(t-2\tau)]}+\ldots. \qquad \ldots\ldots(50)$$

With this assumption the second equation of (46) is independent of the first and can be written

$$\frac{d\eta}{dt}=-\frac{1}{CZ_0}h_2(t). \qquad \ldots\ldots(51)$$

Also it is easy to see that

$$\left.\begin{aligned}H(t-\tau)&=\frac{H(t)}{\rho e^{-i\psi}}-1,\\ \text{where}\qquad \psi&=\omega_0\tau+\eta(t)-\eta(t-\tau).\end{aligned}\right\} \qquad \ldots\ldots(52)$$

Splitting (52) into its real and imaginary parts,

$$\left.\begin{aligned}h_1(t)&=\rho\cos\psi[1+h_1(t-\tau)]+\rho\sin\psi h_2(t-\tau),\\ h_2(t)&=-\rho\sin\psi[1+h_1(t-\tau)]+\rho\cos\psi h_2(t-\tau).\end{aligned}\right\} \qquad \ldots\ldots(53)$$

Equation (51) in conjunction with (53) is not difficult to deal with numerically. If a solution has been determined and a better approximation is desired it should be possible to achieve this by an iterative method applied to (46). The solution can be substituted in the first equation of (46) to obtain a first approximation to $|V|$. This can be used then in the second equation to obtain a second approximation to η and so on. It seems reasonable to assume that this sequence of approximations converges fairly rapidly.

The condition for a steady solution is that $h(t)=h(t-\tau)$, and then from (51) and (52)

$$\frac{d\eta}{dt}=-\frac{1}{CZ_0}\operatorname{Im}\frac{\rho}{e^{i\psi}-\rho}=-\frac{1}{2CZ_0}\frac{\sin\psi}{\xi+\cos\psi}. \qquad \ldots\ldots(54)$$

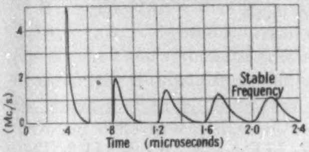
This is, in fact, the condition (21) derived for the static case, and is only strictly valid when $d\eta/dt$ and ψ are constant. However, it has been suggested empirically by Mr. C. L. Ratsey that if we use this equation in the general case we should get quite a useful qualitative idea of the resulting frequency. Numerical solutions of equations (52) and (54) for a typical case, which were integrated by Dr. E. T. Goodwin, did indicate qualitative agreement. Equation (54) also has the advantage that it can be put on a small differential analyser, and solutions for many different values of α, $1/CZ_0$, τ, $\omega_0\tau$ can then be run off. This was done on the small differential analyser at Cambridge for values of

$$\alpha=0{\cdot}5,\ 0{\cdot}68,\ 0{\cdot}95; \qquad \tau=0{\cdot}1,\ 0{\cdot}25,\ 0{\cdot}4\ \mu\text{sec.};$$

$$1/CZ_0=240; \qquad \omega_0\tau=170^\circ,\ 60^\circ,\ 30^\circ.$$

The functions η and ψ were integrated directly and then $d\eta/dt$, or instantaneous frequency, can be determined from (54). Three typical solutions are reproduced in figures 11, 12, 13. As is expected solutions near $\omega_0\tau=0^\circ$ are stable and settle down quickly, whereas those in the neighbourhood of $\omega_0\tau=180^\circ$, where variations in the initial conditions of rise can cause instability, take a long time to attain the stable frequency.

Figure 11



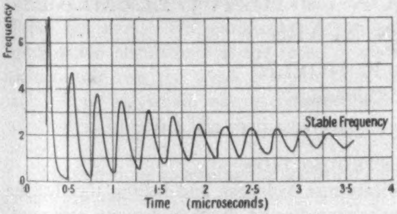


Figure 12.

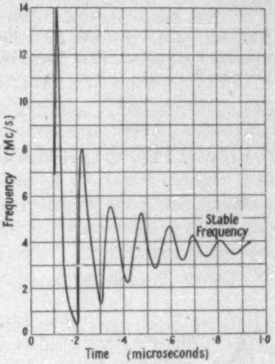


Figure 13.

The experimental results in §2 did not yield $g(|V|)$ directly, but the power output as a function of $|V|\sqrt{C}$, say $g_1(|V|\sqrt{C})$; if we wish to use these results we can still deduce relative amplitudes by the transformation,

$$V_1=V\sqrt{C}, \quad F_1=F\sqrt{C}.$$

We then obtain on substituting in (43)

$$\frac{dV_1}{dt}+\left[\frac{1}{2CZ_0}-\frac{g_1(|V_1|)}{|V_1|^2}\right]V_1=-\frac{\rho}{CZ_0}e^{-i\omega_0\tau}F_1(t-\tau). \qquad \ldots\ldots(55)$$

If we are using the simple formula (37) for $g(V)$ it becomes convenient to express voltages in terms of V_0 as unit. Writing $V/V_0=V_2$, $F/V_0=F_2$ (43) assumes the form

$$\frac{dV_2}{dt}+\left[\frac{1}{2CZ_0}-\frac{1}{2CR_0}\left(\frac{1}{|V_2|}-1\right)\right]V_2=-\frac{\rho}{CZ_0}e^{-i\omega_0\tau}F_2(t-\tau). \qquad \ldots\ldots(56)$$

he instability in the neighbourhood of $\omega_0\tau=0$ and the probability of either solution for any particular value of $\omega_0\tau$ has not been discussed theoretically. It must involve a detailed consideration of the initial conditions of rise of the oscillator, and is a problem of considerable difficulty.

ACKNOWLEDGMENTS

The work described was carried out in the Admiralty Signal Establishment, and the paper is published with the approval of the Board of Admiralty. The author is indebted to the Captain Superintendent for granting facilities for the preparation of the paper. He also wishes to express his thanks to Dr. W. T. Davies for drawing his attention to this problem and for many useful discussions concerning it; to Dr. E. T. Goodwin for his advice and help in connection with the approximations and numerical integrations; to Mr. J. Crank and the staff of the Mathematics Laboratory at Cambridge for their work on the differential analyser; and to Mr. B. W. Lythall for providing experimental data.

REFERENCES

Appleton and Greaves, 1923. *Phil. Mag.*, **45**, 401.

Kryloff and Bogoliouboff (free translation by S. Lefschetz), 1943. *Introduction to Non-Linear Mechanics* (Princeton University Press).

Lythall, 1946. *J. Instn. Elect. Engrs.*, **93**, IIIA, 1081.

van der Pol, 1922. *Phil. Mag.*, **43**, 700.

van der Pol, 1934. *Proc. Inst. Radio Engrs.*, **22**, 1051.

ON THE FORMATION OF HEAVY ELEMENTS IN STARS

By F. HOYLE,
Cambridge

Read 20 *December* 1946 ; *MS. received* 13 *March* 1947

§1. INTRODUCTION

Astrophysical data indicate that hydrogen and helium together are about ten thousand times as abundant as all other elements combined. This statement is based on Russell's general discusssion of the opacity of stellar atmospheres (1933), Strömgren's detailed analysis of the solar atmosphere (1940) and Dunham's estimate for the composition of the interstellar gas (1939). The helium is regarded as arising through synthesis from hydrogen, either by the carbon-nitrogen cycle or by deutron formation (Bethe and Critchfield, 1938; Bethe, 1939). On this basis hydrogen must have been by far the most abundant element during the early history of the universe (it is generally believed that in spite of the subsequent formation of helium, hydrogen is still the most abundant element). It is usually regarded as a natural extrapolation to pass from this conclusion (which is forced on us by observation) to the hypothesis that hydrogen was the only element initially present in the universe. It is then necessary to explain how the heavy elements have been synthesized from hydrogen. The present paper is concerned with a recent attempt to solve this problem (Hoyle, 1947).

It will be useful to divide the discussion into two parts. First we consider the physical conditions necessary for the production of the heavy elements, and second the place in the universe where such conditions are realized will be described. The remainder of the present section, together with §2, will be devoted to the first part, and we shall return to the astronomical considerations in §3.

There are two different ways of approaching the physical part of the problem. We could attempt to synthesize a particular element by specifying in detail a chain of nuclear reactions that start from hydrogen and end with the element in question. Such a method is at first sight attractive, because it can be directly related to processes observed in the laboratory. But a non-statistical method breaks down in attempting to proceed in a reverse direction along the radioactive series beyond such very short-lived nuclei as RaC′ and ThC′. This difficulty does not arise in a statistical treatment, and for this reason alone it would be preferable to adopt the method of statistical mechanics (in which it is not necessary to specify individual processes in detail).

§ 2. THE PHYSICAL PART OF THE PROBLEM

It is well known in the theory of dissociating gases that if for the n substances $A_1, A_2, \ldots A_n$ it is possible to find a chain of reactions that connects A_r with A_s (for all pairs r and s) then the relative abundances of the substances can be determined entirely from statistical mechanics. The statistical treatment assumes that the system is given sufficient time to reach equilibrium. This proviso is very important in the present problem, because it restricts the search for a suitable place where the synthesis of the elements may occur to a particular class of exceptionally dense, hot stars. The astronomical considerations of § 3, taken together with quantitative calculations of the speed of nuclear reactions, show that statistical equilibrium over the whole periodic table requires temperatures greater than about $4 . 10^9$ °c. (which may be compared with a central temperature in the Sun of about $2 . 10^7$ °c.). We shall confine attention to temperatures exceeding this value.

The equations governing the statistical equilibrium between nuclei are closely similar to the equations for dissociating gases (Sterne, 1933; Fowler, 1936). When the density ρ and the temperature T are given, the equations determine the abundances of the elements. Thus for each pair of values of ρ, T there is a unique composition for material under statistical equilibrium. If we regard $\log_{10}\rho$ (ρ in gm. per cm.^{3}) and T (in units of 10^9 °c.) as rectangular Cartesian coordinates, there will be a definite composition for material in statistical equilibrium at a given point in the ρ, T plane.

The first general result given by the statistical equations is that the ρ, T plane can be divided into two parts by the curve AEFB of figure 1, which is such that to the right of the curve the material is almost entirely composed of helium, while to the left of the curve the material is largely composed of elements with atomic weights greater than 50.

The composition differs appreciably from one part of the heavy element zone to another. As an example the following table gives the composition at D, which is the point $\rho = 10^7$ gr. per cm.^{3}, $T = 4 . 10^9$ °c. :—

Table 1

Element	^{4}He	^{8}O	^{28}Si	^{56}Fe	^{63}Cu	^{82}Kr	^{116}Sn	^{208}Pb
Logarithm to base 10 of number of nuclei per cm^2	27·2	20·2	23·7	28	26·7	18·8	−3·3	−222

The negligible abundances of elements in the upper half of the periodic table is a striking feature of these values. This result is characteristic of points near D. There is no hope of obtaining the heaviest elements from the region of the ρ, T plane in the neighbourhood of D. For this purpose it is necessary to discuss the composition of material at much higher densities.

Increasing density produces the following important effects:—

(1) A rapidly increasing density of free neutrons.

(2) A rapidly increasing ratio of free neutron density to free proton density.

The increasing ratio of free neutrons to free protons has the effect of increasing the abundances of nuclei containing appreciably more neutrons than protons.

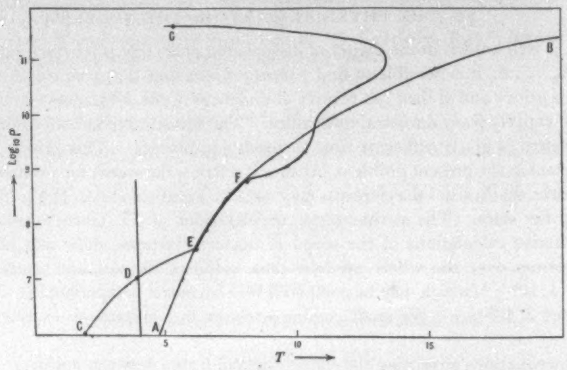


Figure 1. T in units of 10^9 °C. ρ in gr. per cm^3

That is, the abundances of elements at the upper end of the periodic table are increased at the expense of elements in the lower half of the table. This effect is illustrated by the following values:—

Table 2

$\log_{10}\rho'$	11·3	11·5	11·6	11·8
T	9·8	11·6	12·5	14·5
Atomic weight of the most abundant element	130	160	180	240

The quantity ρ' is the difference between the total density ρ of the material and the density of free neutrons (which is comparable with, but less than, ρ at these very high densities). Although it is outside the scope of the present paper to discuss the calculations leading to the values in table 2, it may be noted that the energy of relativistically degenerate electrons is the main factor that controls the properties of matter at very high densities.

The above discussion may be summarized by noting that in the neighbourhood of the point D of figure 1 the composition of material is almost entirely restricted to the lower half of the periodic table, whereas in the neighbourhood of G the composition is mainly confined to the upper half of the periodic table.

§3. THE ASTRONOMICAL PART OF THE PROBLEM

The discussion of §2 narrows considerably the search for places where the synthesis of heavy elements can occur. Normal stars, in which energy production through the conversion of hydrogen to helium balances the loss of radiation at the surface, are known from the theory of stellar constitution to possess central temperatures that are lower than the present requirements by a factor of about a hundred.

The Sun possesses hydrogen sufficient to maintain the present normal state for a time of about 10^{11} years, which is considerably greater than the usual estimates for the age of the universe. Accordingly there is no prospect of a departure from the normal state in small stars such as the Sun. The situation is very different, however, for massive stars. The energy production necessary to balance the loss of radiation from the surface is so large in a star of twenty solar masses, for example, that the supply of hydrogen becomes exhausted in about 10^8 years (unless further supplies of hydrogen are added by a rapid sweeping up of the interstellar gas). If the hydrogen should become exhausted, energy production by the carbon-nitrogen cycle, and by any other process involving proton reactions, ceases. On the other hand the loss of energy by radiation at the surface *must* continue as before. This loss of energy then leads to a slow contraction of the star. Now it is well known in the theory of the constitution of the stars that the central temperature of a *slowly* collapsing star is approximately inversely proportional to the radius. Thus the central temperature rises as the radius decreases, and it is clear that if sufficient contraction occurs in massive stars that have exhausted their internal supply of hydrogen, then temperatures of order $5 . 10^9$ °c. must be attained.

By good fortune it is possible to make a quantitative test of this suggestion. From the work of Baade (1942) and Minkowski (1942) on the supernova of A.D. 1054 we can obtain estimates for the mass and radius of the star before the supernova outburst occurred. These values are about fifteen solar masses, and about one twentieth of the solar radius, respectively. Since the central temperature is proportional to the ratio of mass to radius, its value in this case must be about 300 times that of the Sun, or about $6 . 10^9$ °c., which is in excellent agreement with our requirements.

We next consider the properties of collapsing stars in further detail. Chandrasakhar has shown that if the mass of a contracting star exceeds about 1·5 solar masses, then the star cannot attain a spherically symmetric equilibrium state, no matter how long contraction continues. This means that contraction would continue in a non-rotating star until the radius became comparable with the gravitational radius (which is only a few kilometres, even for massive stars). On the other hand, in a rotating star, centrifugal forces become important as the contraction continues, and a stage must be reached at which the rotational energy of the star becomes so large that its shape becomes markedly spheroidal. In the equatorial plane further contraction is prevented, but contraction perpendicular to this plane will continue if the star is sufficiently massive until the polar radius is about half the equatorial radius. As Jeans showed, the star now becomes unstable, and a sharp edge develops in the equatorial plane, through which material is ejected. The motion of this material is very difficult to trace by theoretical analysis. It is reasonable, however, to assume that the material behaves in accordance with the observed motions of material ejected by novae and supernovae.

The degree of contraction required to produce instability depends on the amount of angular momentum possessed by the star, and this is known to vary widely from one star to another. Thus if a star possesses a large amount of angular momentum (corresponding to an equatorial rotational velocity in the normal state of several hundred km. per sec.), the degree of contraction required to induce rotational instability is small, and the velocity of ejection of material is small (several hundred km. per sec.). On the other hand, if we take a massive star with the same angular momentum per unit mass as the Sun, for example, the degree of contraction is very large indeed, and the velocity of ejection is correspondingly large (several thousand km. per sec.). So we have a whole range of cases from mild to extremely violent outbursts, according to the angular momentum of the star. It seems possible to fit a variety of observed phenomena such as Wolf-Rayet emission, novae and supernovae in this range.

§ 4. THE SYNTHESIS OF THE HEAVY ELEMENTS

We can now describe the changes that occur in the composition of material in a collapsing star. The curve CDEFG of figure 1 is a representative case worked out from the theory; it has been shown that material near the centre of a collapsing star of five solar masses must evolve along this track. Before reaching D the material is mainly composed of helium. Nuclear reactions become sufficiently rapid for statistical equilibrium to be attained in the neighbourhood of D, and between D and E the helium is converted into a distribution of the form shown in table 1.

This synthesis of heavy elements from helium yields energy, which is then available to balance the loss of radiation at the surface. Accordingly the collapse of the star is temporarily arrested by this energy production, but is resumed when the conversion of the helium is completed. The delay in the contraction may last as long as a million years. This delay is in striking contrast with the behaviour of the star after the point E has been reached. For after passing E the internal pressure is inadequate to provide even approximate support of the star against gravity. An extremely rapid collapse ensues, leading to an evolution from E to G in about a hundred seconds. The reason for this remarkable behaviour is readily understood. For it is clear that the direction of the tangent to the curve must change discontinuously at E, since no energy is available at E to reconvert into helium the heavy elements synthesized between D and E. Such a reconversion requires a supply of energy that can only be forthcoming from a rapid collapse of the star. During this contraction the material follows a section of the curve bounding the helium zone. Thus during the evolution along the section EF of the track, the star is supplying gravitational energy which is absorbed in the reconversion of heavy elements to helium. The reconversion is completed at F, and the material can now enter the helium zone. It might seem that the star can now return to a state of approximate mechanical equilibrium. This is not the case, however, on account of the increasing importance of free neutron production which leads to a further absorption of large quantities of energy. The form of the track between F and G is due to the effect of free neutron production taken together with the non-linear properties of the equations of statistical equilibrium. The sharp decrease of temperature in the neighbourhood of G slows down the rate of the nuclear reactions to such an extent that it is doubtful whether the statistical equations are adequate to describe the form of the track beyond G.

It is tacitly assumed in the above discussion that the evolution of material along CDEFG is not interrupted by the onset of rotational instability. The remarks of the previous section show that the stage at which the evolution is interrupted depends on the angular momentum of the star. If the angular momentum is large enough, instability will occur before material at the centre of the star has evolved beyond E. In this case the synthesis of heavy elements is confined to the lower half of the periodic table. On the other hand, in stars with sufficiently small angular momentum, the material can reach the neighbourhood of G before the onset of instability. The slowing down of the rate of nuclear reactions near G is strongly indicative that in this case the composition of material will remain frozen during the instability process.

It is seen, therefore, that elements in the lower half of the periodic table are provided by material near D, whereas the elements in the upper half of the table require the material to reach the neighbourhood of G before the instability occurs. In this connection it may be noted that the central density of a star is about thirty times the mean density. Consequently there must be a considerable variation in the composition of material in different parts of a rotationally unstable star.

§ 5. GENERAL REMARKS

The theory described above, when taken together with the hypothesis that hydrogen was the only element initially present in the universe requires the first stars to be initially composed of pure hydrogen. It is known from the study of stellar structure that the time required for a massive star to exhaust its supply of hydrogen is substantially independent of the mode of conversion of the hydrogen to helium. Thus about 10^8 years after the formation of the first massive stars, it is to be expected that heavy elements would begin to be distributed in interstellar space. Subsequent stellar condensations are formed from the initial hydrogen together with these heavy elements. It follows, therefore, that pure hydrogen stars condense only in the earliest stages in the evolution of a galaxy. The relation of this question to the origin of the white dwarf stars has been discussed in a recent paper (Hoyle, 1947 b).

A star that becomes rotationally unstable after the central material reaches the point E in the ρ, T plane must undergo an extremely violent outburst, since the unstable state is reached in a catastrophic manner. On the other hand, a star that becomes unstable before the central material arrives at E takes many thousands of years to reach the unstable state. The critical nature of the point E provides a natural explanation of the difference between novae and supernovae. That is, supernovae are collapsed stars that become unstable *after* the point E has been reached, whereas novae are collapsed stars that reach the unstable state *before* the point E is attained.

It remains to show that the present theory is capable of supplying the required quantity of heavy elements. The average rate at which elements are distributed in interstellar space is evidently of the same order as the mean rate at which material is ejected by supernovae. Using the observational estimate of one supernova per galaxy per 500 years we obtain, on the basis of the ejection of ten solar masses per supernova, about $2 \,.\, 10^8$ solar masses distributed as heavy elements in interstellar space in 10^{10} years (this time is in accordance with the usual estimates for the age of the universe). This gives an abundance of heavy elements amounting to about 0·1% by mass of the hydrogen abundance. The remarks made at the outset, concerning the relative abundance of heavy elements and hydrogen, show that this estimate is of the required magnitude.

REFERENCES

BAADE, W., 1942. *Astrophys. J.*, **96,** 188.
BETHE, H. A., 1939. *Phys. Rev.*, **55,** 434.
BETHE, H. A., and CRITCHFIELD, C. L., 1938. *Phys. Rev.*, **54,** 248.
DUNHAM, T., Jr., 1939. *Proc. Amer. Phil. Soc.*, **81,** 277.
FOWLER, R. H., 1936. *Statistical Mechanics* (Cambridge : The University Press), p. 655.
HOYLE, F., 1947 a. *Mon. Not. R. Astr. Soc.* (in course of publication).
HOYLE, F., 1947 b. *Mon. Not. R. Astr. Soc.* (in course of publication).
MINKOWSKI, R., 1942. *Ap. J.*, **96,** 199.
STERNE, T. E., 1933. *Mon. Not. R. Astr. Soc.*, **93,** 736.
STRÖMGREN, B., 1940. *Festschrift für Elis Strömgren.*
RUSSELL, H. N., 1933. *Astrophys. J.*, **78,** 239.

DISCUSSION

Dr. MARTIN JOHNSON. The following difficulties occur to me: All known stars exhibit hydrogen in considerable abundance, and yet Dr. Hoyle's mechanism can only begin when all hydrogen has been exhausted. Heavy elements are also observed in the same stars though their genesis ought to have come via Dr. Hoyle's mechanism. It seems likely that most stars entering on his mechanism of collapse would be instantly re-expanded by their remaining hydrogen before his densities and temperatures are reached.

Prof. R. E. PEIERLS. One should bear in mind that in non-equilibrium conditions, the formation of heavy elements may be possible at pressures and temperatures at which the amount of such elements in thermal equilibrium would be negligible. To use an analogy, the distribution over the surface of the earth of organic compounds probably does not correspond to the abundances they would have at any temperature or pressure. If, in the course of reactions between light elements, free neutrons are produced, they may lead to a gradual building up of heavier elements.

Most of the discussions of abundance seem to start from the hypothesis that the primary element is hydrogen and in some stage the universe consisted of hydrogen only. This is the simplest, but by no means the only, possible hypothesis, and if we explain the origin of heavy elements by their formation from hydrogen the next question is evidently, where does the hydrogen come from?

AUTHOR'S reply. There is no difficulty in understanding how both hydrogen and heavy elements may come to be present in the same star. All that is required is for the star to condense *after* the heavy elements began to be distributed in interstellar space. Moreover, even if a star were initially composed of pure hydrogen it could still acquire heavy elements through the subsequent accretion of interstellar material.

Stars such as V Pupp. and Y Cygn. must exhaust their internal supply of hydrogen in about 10^8 years, unless they are replenished with hydrogen by accretion. Thus in a time that is, astronomically speaking, very short, a number of very bright stars, now observed in a normal state, must evolve along the collapsing sequence that leads to the processes I have described. The situation is even more marked in the case of 29 Can. Maj. and Ao Cass., where the hydrogen supply, if unreplenished, will be exhausted in about 10^7 years.

In using the analogy of organic compounds it must be remembered that the high-grade radiation from the Sun plays an essential part in the building of these compounds. The presence of a corresponding high-energy source would be necessary in a non-statistical approach to the present problem. Even if such a source were forthcoming I think great difficulty would still arise in attempting to pass the barriers at RaC′ and ThC′.

The observed high abundance of hydrogen is strongly suggestive that only hydrogen was initially present in the universe. The question of the origin of the hydrogen may well be of great significance, but this problem must be left over for future discussion.

NUCLEAR MAGNETIC MOMENTS

BY R. LATHAM,
Cavendish Laboratory, Cambridge

MS received 20 *March* 1947

ABSTRACT. A regular change in the nuclear magnetic moments of the elements as a function of their charge and spin is described, and a connection is suggested between this and the excess of the number of neutrons over the number of protons in the nucleus.

§ 1. INTRODUCTION

THE purpose of this note is to point out an empirical regularity of behaviour in the nuclear magnetic moments of the elements as a function of their charge and spin. The kind of correlation found would require more extensive experimental data to establish it on a firm basis, and it must therefore be regarded as tentative.

The nuclei, whose spins and magnetic moments are listed in various tables (Breiglеb, 1940; Mattauch, 1942; Kellogg and Millman, 1946), can be divided into four types dependent on whether the number of protons (p) and neutrons (n) they contain is even or odd. It is then found, as is well known, that the spins and magnetic moments are different for the different types as shown in table 1.

Table 1

Nuclear type	No. of protons and neutrons		Spin	Value of moment (nuclear magnetons)	Dependence on spin
A	$2p$	$2n$	0*	0*	—
B	$2p$	$2n+1$	half integral	-1 to $+1$ approx.	no effect
C	$2p+1$	$2n$	half integral	$-\frac{1}{2}$ to $+6$ approx.	increases with spin
D†	$2p+1$	$2n+1$	integral	$+0{\cdot}4$ to $+0{\cdot}9$ approx. for light nuclei	—

§ 2. EMPIRICAL RELATIONS

Plots of the nuclear magnetic moment against spin for nuclei of types B and C have previously been made, as for example, by Way (1932), Wigner and Feenberg (1941) and Margenau and Wigner (1940), the results being as shown in figures 1 and 2. It will be seen that for each type the points lie near to two parallel curves, though not exactly on them. The curves of course have no existence between the integral or half integral values of the spin coordinate but are useful in displaying the variation with spin. Nuclei with an odd neutron (B) give curves staying more or less horizontal, while for nuclei with an odd proton (C) the curves rise with increasing spin. The points for very light nuclei of $Z \leqslant 9$ do not lie on the curves and have not been plotted, but all these omitted points are given as tables on the two figures. The deviations from the curves for the heavier nuclei are in many cases well outside experimental error but rarely large enough to be appreciable compared with the separation of the curves.

---

* Either zero or presumed zero because of absence of hyperfine structure.

† Only four light stable nuclei of this type are known, and two heavier nuclei $^{40}_{19}$K and $^{176}_{71}$Cp.

If now the nuclei are displayed on a diagram whose coordinates are spin and nuclear charge Z (or number of protons) it is seen (figure 3) that there are regions within which all the points belong either to the higher or to the lower curve. The boundaries are not well defined due to the small number of points available, but the general effect is clear.

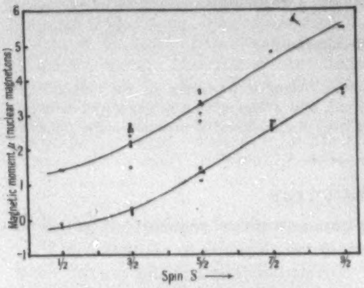


Figure 1. Nuclei of type C ($2p+1$, $2n$).
Points omitted.

	Spin	μ
${}^{1}_{1}$H	1/2	2·789
${}^{7}_{3}$L	3/2	3·253
${}^{11}_{5}$B	3/2	2·686
${}^{15}_{7}$N	(?)1/2	−0·280
${}^{19}_{9}$F	1/2	2·625

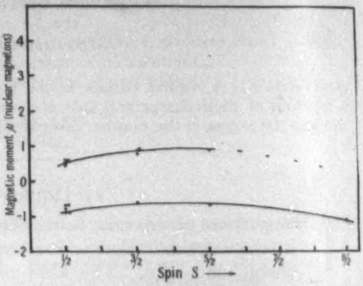


Figure 2. Nuclei of type B ($2p$, $2n+1$).
Points omitted.

	Spin	μ
${}^{1}_{0}$N	1/2	−1·935
${}^{9}_{4}$Be	(?)3/2	−1·176
${}^{3}_{6}$C	1/2	+0·701

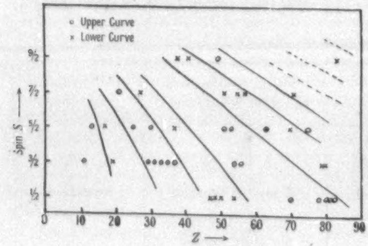


Figure 3.

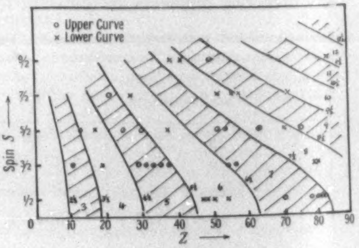


Figure 4.

In trying to express the limits of the regions mathematically some choice must be made, and the following method was adopted. Multiplication of Z by a factor varying with S allows the regions to be made into vertical bands. The appropriate factors are found to be proportional to $(1+\alpha S^2)$ where α is 0·071 as shown in table 2. The limits then occur at values of the quantity $Z(1+\alpha S^2)$ given in table 3, where thay are compared with $1{\cdot}50(k+\frac{1}{2})^2$, k being an integer varying from 3 to 9.

The two factors together would give lines of demarcation dependent on $\frac{1{\cdot}50\,(k+\frac{1}{2})^2}{1+0{\cdot}071\,S^2}$, and these are superimposed on the experimental data in figure 4. It must of course be pointed out that neither the form of the variation with S nor that with k is unique, for example, a dependence on $Z(1-0{\cdot}122\,S)$ would also allow for the effect of spin. The point at $Z=63$ is due to $^{151}_{63}$Eu and $^{153}_{63}$Eu, which have a spin of 5/2 and lie on the upper and lower curves respectively.

Table 2

Spin	$(1+0{\cdot}071S^2)$
0	1·00
1/2	1·02
3/2	1·16
5/2	1·44
7/2	1·87
9/2	2·44

Table 3

$Z(1+\alpha S^2)$	k	$1{\cdot}50(k+\frac{1}{2})^2$
≃ 19	3	18
≃ 29	4	30
47	5	45
≃ 60	6	63
≃ 85	7	84
105	8	108
≃125	9	135

These results suggest that the nuclear magnetic moment for zero spin lies on the upper or lower curve if $c\sqrt{Z}$ (c being a constant) is nearest to an odd or even integer respectively, the odd number giving the higher moment. On the earlier theories of nuclear structure (Bethe and Bacher, 1936), some such dependence on a quantum number is not unlikely. It is also possible that if the spin is partly due to the rotation of the nucleus it may modify the number of particles in a given level.

Figure 5 shows the neutron excess (A-$2Z$) plotted against Z for naturally occurring nuclei of type A, where the spin is probably zero, and below on the same diagram the values of Z for $S=0$, for which the magnetic moment is changing over. The correlation suggests that a lower neutron excess occurs when the change-over is taking place.

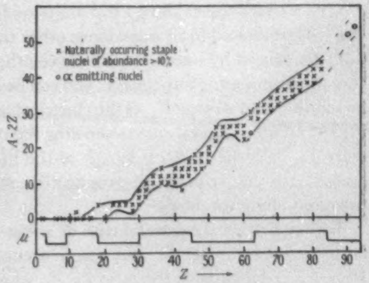


Figure 5.

It would be unwise at present to base much on the degree of agreement shown in figure 4, but if the division into regions could be well established and their boundaries defined more exactly, a useful indication might be obtained as to the kind of structure which exists in the heavier nuclei.

REFERENCES

Bethe and Bacher, 1936. *Rev. Mod. Phys.*, **8**, 171.
Breigleb, G., 1940. *Atome und Ionen* (Leipzig).
Kellogg, G. and Millman, 1946. *Rev. Mod. Phys.*, **18**, 348.
Margenau and Wigner, 1940. *Phys. Rev.*, **58**, 103.
Mattauch, J., 1942. *Kernphysicalische Tabellen.*
Way, K., 1939. *Phys. Rev.*, **55**, 963.
Wigner and Feenberg, 1941. *Reports on Progress in Physics (Phys. Soc.)*, **8**, 294.

THE MATCHING OF HIGH-FREQUENCY TRANSMISSION LINES USING A FREQUENCY-VARIATION METHOD

By A. S. EDMONDSON,
The Inter-Services Research Bureau;
now at the Sir John Cass Technical Institute, London

MS. received 12 *March* 1947

ABSTRACT. A frequency-variation method which shows when an impedance is correctly matched to a transmission line is described. The variation with frequency in the sending-end voltage of a transmission line many wavelengths long is reproduced on the screen of a cathode-ray oscilloscope and from the curve it is possible to see whether the standing waves along the line are large or small in amplitude, and how they vary with frequency. A description is given of an apparatus applying the new method to ultra-high frequencies and used in connection with the matching of aerials and filters to a coaxial transmission line. The method has also been used for demonstrating several important transmission line properties and in this connection is useful for educational purposes.

§1. THE FREQUENCY-VARIATION METHOD

STANDING waves along a transmission line of finite length occur when the line is terminated in an impedance other than its characteristic impedance; they are caused by interference between the incident wave and the wave reflected from the terminating impedance, and can be detected by measuring the variation in the sending-end voltage V_s, as the electrical length βl of the line is altered. Neglecting attenuation, typical curves showing the relation between V_s and βl are shown in figure 1. (l is the physical length of the line, $\beta = 2\pi/\lambda = 2\pi f/v$, where f = the frequency, v = the velocity of propagation along the line and λ = the wavelength measured along the line.)

Figure 1 (*a*) is the general case of a loss-free line terminated by an impedance Z_T having both resistive and reactive components. The standing-wave ratio ($V_{\max}/V_{\min}$) depends on the ratio Z_T/Z_0, where Z_0 is the characteristic impedance of the line. Figure 1 (*b*) is the case of the line open- or short-circuited, or terminated by a reactance only; the shapes of the curves are the same for each, but the distances of the maxima and minima from the end of the line are different. The standing-wave ratio is theoretically infinite. Figure 1 (*c*) shows the relation when the line is terminated by its characteristic impedance; there are no standing waves and the voltage is independent of the electrical length of the line. In all practical cases, attenuation causes the incident and reflected waves to decrease in amplitude exponentially with distance along the line, resulting in a diminution of the standing-wave ratio as the distance from the termination increases.

The electrical length of the line cannot be altered satisfactorily by decreasing l owing to the practical difficulty of increasing it again; it is therefore changed by varying the frequency. It is shown later that the line must be long in order that the curve showing the variation of voltage at the sending end with frequency, the (V_s, f) curve, should have several maxima and minima when standing waves are present. The general type of curve obtained when the line is terminated by an impedance Z_T is shown in figure 2(*a*). The exact positions of the maxima, as well as their number, depend on the length of the line; the broken line shows the envelope of the series of curves that would be obtained if l were increased or decreased gradually by an amount $\lambda/2$. When Z_T is constant, the general shape of the curve is the same over the frequency range, but when Z_T varies, the relative amplitudes of the voltage maxima alter accordingly. If the range of frequencies includes a region in which the value of Z_T approaches the characteristic impedance of the line, the envelope of the (V_s, f) curve shows a minimum as in figure 2(*b*), since then the standing-wave ratio is least. If Z_T becomes equal to Z_0, the range of frequencies for which this condition is fulfilled gives no variation of V_s and the envelope of the curve crosses over where matching occurs, as in figure 2(*c*).

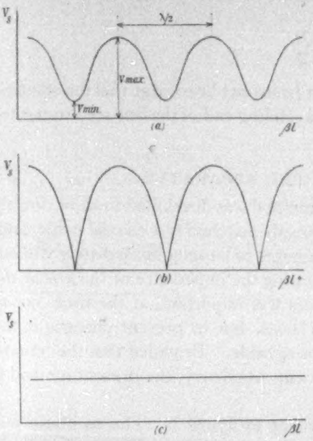


Figure 1. The variation of the sending-end voltage V_s of a transmission line with the electrical length βl;
(*a*) when the line is terminated in an impedance $Z_T = R_T + jX_T \neq Z_0$;
(*b*) when the line is terminated in an open- or short-circuit or a pure reactance;
(*c*) when the line is terminated in its characteristic impedance Z_0.

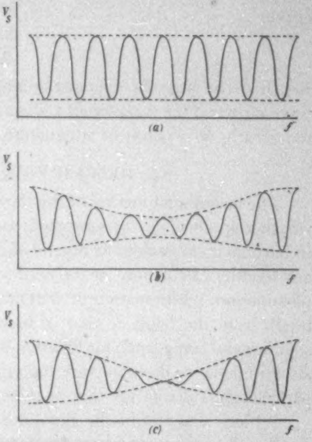


Figure 2. The variation of the sending-end voltage V_s of a transmission line, many wavelengths long, with frequency f;
(*a*) when $Z_T \neq Z_0$ and does not vary appreciably with frequency;
(*b*) when Z_T approaches Z_0 at a frequency within the range of observation;
(*c*) when $Z_T = Z_0$ at a frequency within the range of observation.

So that changes in the standing waves along the line due to a small change of Z_T may be readily detected on the (V_s, f) curves, the maxima must be closely spaced, i.e. $\Delta N/\Delta f$ should be as large as possible, ΔN being the number of waves appearing in the (V_s, f) curve in the frequency range Δf. Since the maxima of the standing waves along the line are spaced at intervals of $\lambda/2$, the number of half-wavelengths, N, in the line measures the number of standing waves at a frequency f, i.e.

$$N=\frac{l}{\lambda/2}=\frac{2lf}{v}.$$

The change in frequency Δf producing a change in N of ΔN is given by the equation

$$\Delta N=\frac{2l}{v}\Delta f.$$

But since ΔN is also the number of waves in the frequency range Δf on the (V_s, f) curve,

$$\frac{\Delta N}{\Delta f}=\frac{2l}{v}.$$

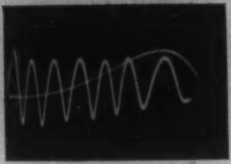
For this to be large the line must be long, but l must not be so large that the standing wave ratio and the variation of voltage at the sending end of the line are diminished too greatly on account of attenuation.

§ 2. DESCRIPTION OF THE APPARATUS

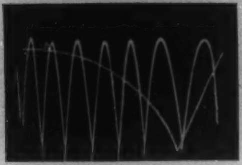
The instrument incorporating the new method was developed to show directly whether or not a piece of equipment was correctly matched to a coaxial cable, and, in particular, to enable the results of adjustments to be seen immediately without any lengthy calculations or the need for knowing the impedance of the line or the termination. Elimination of standing waves was important, at the time, not so much from the point of view of reducing losses, but to prevent changes in the sending-end impedance for different lengths of cable. Provided that the terminating impedance does not vary too rapidly with frequency, the present method is particularly suitable for this purpose.

A block diagram of the apparatus is shown in figure 3. The oscillator was fitted with a rotary condenser driven by a synchronous motor to vary the frequency. The output of the oscillator was taken to the long transmission line, connected at its far end to the terminating impedance, while the high-frequency voltage variations at the sending end of the line were rectified by a detector unit. The low-frequency voltage variations from the detector were amplified and applied to the Y-plates of the cathode-ray oscilloscope. The linear timebase of the oscilloscope was synchronized to the rotary condenser through a phase and amplitude control unit.

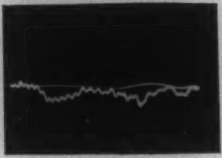
The oscillator was driven by two triode valves working in push-pull, and was tuned by a $\lambda/4$ resonant line short-circuited by a sliding bar which altered the mean frequency from about 280 to 400 megacycles per second. For a fixed setting of the shorting bar the oscillator was tuned over a range of about 15 megacycles per second by the rotary condenser, which was of the series-gap type and altered the capacity at the open end of the resonant line. Variable resistors in the anode and grid circuits enabled the oscillator to be operated under suitable conditions. The frequency of the ocillator was determined mainly by the position of the shorting bar and the setting of the variable condenser, but was affected slightly by the loading. The exact frequency for a given set of conditions was measured with an absorption wavemeter employing a tuned coaxial line resonator and a silicon-tungsten crystal, with a sensitive meter to indicate resonance.



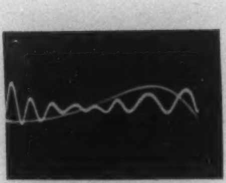
No. 1.



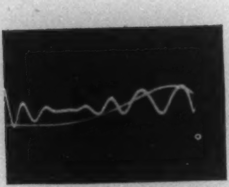
No. 2.



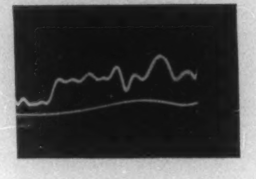
No. 3.



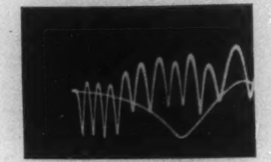
No. 4.



No. 5.



No. 6.



No. 7.

The transmission line was a coaxial cable (Telcon PT29M), consisting of a single inner conductor of copper insulated by polythene from the outer conductor of copper braiding. The characteristic impedance was 72 ohms and the attenuation, at 300 megacycles per second, was about 0·13 decibels per metre. The ratio (wavelength along the cable)/(wavelength in air) had a value approximately equal to 2/3. The most suitable length of cable depended on the type of experiment being performed, but about 50 metres, for which $\Delta N/\Delta f = 0{\cdot}5$ cycles per Mc./sec., was satisfactory for many applications in the present frequency range.

When a constant impedance load equal to the characteristic impedance of the cable was required a very great length was used, since, owing to the attenuation of the reflected wave, the standing-wave ratio at the sending end was small even when the line was not correctly terminated. Z_s therefore remained substantially constant and equal to Z_0 over the whole frequency range. In practice, using a cable 185 metres long terminated in a resistance approximately equal to Z_0, the variation in voltage observed at the sending end as the frequency was altered was very small.

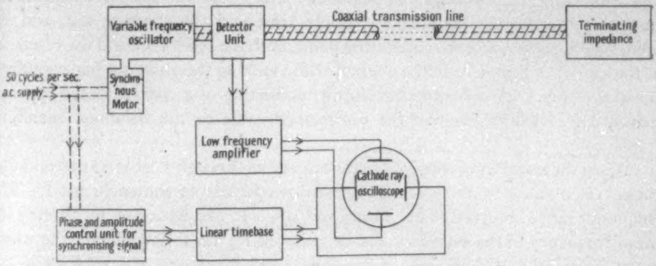

Variable frequency oscillator
Detector Unit
Coaxial transmission line
Terminating impedance
50 cycles per sec. a.c. supply
Synchronous Motor
Low frequency amplifier
Cathode ray oscilloscope
Phase and amplitude control unit for synchronising signal
Linear timebase

Figure 3. Block diagram of the apparatus.

In using the apparatus qualitatively, i.e. simply to detect the presence of standing waves or to compare their relative magnitudes under different conditions, variations only of the high frequency voltage needed to be recorded; precautions necessary to measure the absolute voltages at these frequencies were therefore not required. The rectifier, using either a crystal or a diode valve, was built into a detector unit, which, to measure the voltage variations, was connected across the cable at the sending end. The output was applied to the Y-plates of the oscilloscope through the low-frequency amplifier. The curve obtained on the screen showed the variation of the sending-end voltage of the line with frequency.

§ 3. EXPERIMENTAL RESULTS

Photographs of the oscilloscope traces obtained in different cases are shown in nos. 1–7 of the plate.

No. 1 shows the general case of a cable, 43 metres long, terminated in an impedance Z_T. The voltage variation is similar to that shown in figure 2 (*a*).

No. 2 is the oscillogram for the same length of cable short-circuited at its termination. It has the same shape as the curve in figure 1 (*b*) and shows that the sending-end voltage varies over comparatively wide limits.

No. 3 was obtained using a cable 185 metres long terminated in a resistance approximately equal to Z_0. The small variations in voltage due to the very great length are seen, showing that the sending-end voltage of the line remained fairly constant. The larger variations were due to irregularities in the cable which are explained later; because of these irregularities, this sample of cable could not be used as a constant load for experimental purposes, but, at the time the photographs were taken, it was the only really long piece available.

Nos. 4 to 7 are explained in the following applications which demonstrate the usefulness of the new method.

Matching a dipole aerial to a cable

A dipole aerial, specially made for demonstration purposes, was constructed so that both the resistive and reactive components of its impedance could be adjusted over wide ranges, thus enabling it to be matched to the cable at any frequency within the range of the oscillator. The dipole arms were fixed at one end of a pair of parallel lines acting as a matching transformer, while at the other end was a shorting bar, adjustable in position. The cable was connected across the matching transformer between the dipole arms and the shorting bar, and, by adjusting the position of the connecting point, both the resistance and the reactance at the end of the cable could be altered, while moving the shorting bar varied the reactance only. (A balancing transformer consisting of a coaxial sleeve approximately $\lambda/4$ in length coupled the unbalanced cable to the balanced matching transformer.)

When the aerial was connected to the apparatus through a cable 43 metres long, the curve obtained on the screen of the oscilloscope was as shown in no. 1. The frequency range observed at any instant was about 15 Mc./sec., but, by altering the mean frequency of the oscillator, the variation of V_s was followed over the whole range from 280 to 400 Mc./sec.

The aerial was matched to the cable at a particular frequency in the following way. The required frequency was brought to a known point on the screen by adjusting the mean frequency of the oscillator, and the positions of the cable connecting point and the shorting bar on the matching transformer were then adjusted until a minimum in the voltage variation was observed as in no. 4. It was found that this could soon be done, and no difficulty was experienced in obtaining the necessary conditions. By means of further small adjustments a curve was obtained having a flat portion, indicating the disappearance of standing waves along the cable and showing that matching occurred at that point. The point at which matching occurred was then brought to the required position on the frequency axis by making suitable adjustments to the aerial. No. 5 shows the matching curve finally obtained.

The effect of reflection of the radiation from the aerial by a neighbouring object was observed by the change in the curve on the oscilloscope screen. A reflector was brought slowly up to the arms of the aerial originally matched to the cable feeding it. The changes in impedance so caused had an effect on the standing waves along the cable, and periodic variations in the (V_s, f) curve were observed, which increased greatly in magnitude as the reflector approached the aerial. For a given position of the reflector, in order to match the aerial to the cable again, further adjustments of the cable connecting point and the matching transformer shorting bar had to be made.

Adjustment of ultra-high frequency filters

In a particular application it was required to short-circuit one frequency f_1 without causing any mismatch to a second frequency f_2. If λ_1 and λ_2 are the wavelengths corresponding to the frequencies f_1 and f_2 respectively, an open-circuited stub of electrical length $\lambda_1/4$ will short-circuit f_1 when placed across a line. The reactance of this stub at the frequency f_2 can be tuned out with a second open-circuited stub whose electrical length is $(\lambda_2/2-\lambda_1/4)$ without affecting its short-circuiting properties at the first frequency. Figure 4 (*a*) shows the arrangement of the stubs across the line.

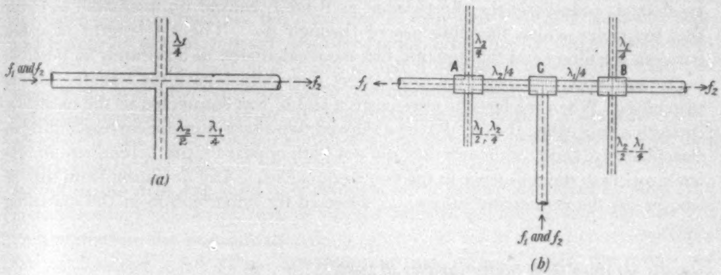


Figure 4. High-frequency filters.
(*a*) Filter short-circuiting f_1 and matched to f_2.
(*b*) Filter separating two frequencies f_1 and f_2.

A junction box was constructed, having two coaxial side-plugs for connecting the stubs across a coaxial cable. The stubs themselves consisted of short lengths of coaxial cable attached to sockets which fitted on to the side-plugs of the junction box. The electrical length of a stub could be altered by shortening, but for experimental purposes it was more convenient to connect a small vaaiable condenser to the open end to give a continuous adjustment over a small range.

To one end of the junction box was connected the usual long piece of cable leading to the oscillator and to the other end a very long piece of the same type of cable was connected to provide a constant impedance load. To short-circuit the first frequency, a stub was fitted to one side-plug, and a detector unit, whose amplified output was applied to the Y-plates of the oscilloscope, was fitted to the other. The curve obtained on the screen showed the variation with frequency of the voltage across the line at the junction; the frequency at which a short-circuit was produced was shown by a pronounced minimum on the curve. The electrical length of the stub was adjusted until the minimum occurred at the required frequency f_1. To tune out the reactance of the first stub at the second frequency, the detector unit was placed at the oscillator end of the cable, to measure V_s, and a second stub was fitted to the second side-plug. The output of the detector unit was applied to the oscilloscope and the usual matching curve was obtained. By adjusting the electrical length of the second stub, matching was achieved at the frequency f_2. In contrast to the relatively wide flat portion on the matching curve in the case of the dipole aerial, matching occurred at a single well-defined frequency in the case of a filter, and the stub lengths had to be correct to within close limits.

A further problem in connection with filters, which was solved with the aid of the new method, was the separation of two frequencies f_1 and f_2 without causing any mismatch to either. One filter, A, similar to that described above, was adjusted so that it was matched to f_1 while short-circuiting f_2, and another filter, B, was matched to f_2 while acting as a short-circuit to f_1. A was connected to a junction box C, see figure 4 (*b*), through a cable of electrical length $\lambda_2/4$, and B was connected to C through a $\lambda_1/4$ length of cable. The short-circuit at A appeared at C as an open-circuit, thus causing no mismatch to f_2 which passed through B unaffected. Similarly the short-circuit at B for f_1 caused no mismatch at C for that frequency which therefore passed through A. The $\lambda/4$ lengths of cable between the filters and the junction box were calculated as accurately as possible before assembly, but their final values were those which gave the best results for matching. When the lengths were correct and C was connected to the oscillator through a long cable, the (V_s, f) curve showed two sharp minima corresponding to matching at f_1 and f_2, although they did not both appear on the screen simultaneously owing to the difference in the two frequencies. Any deviation from the $\lambda/4$ lengths for the connecting cables was detected by imperfection in the matching curve.

Results obtained with faulty samples of cable

With certain samples of cable the usual flat portion of the matching curve was not obtained, even though it was known that the line was correctly terminated. There was a general flattening of the curve but it was distorted as though the usual curve were superimposed upon another. An example of this is shown in no. 6, which was the closest approach to the usual matching curve that could be obtained in a particular case. No. 7 shows the same cable, 76 metres long, short-circuited at its end. These photographs should be compared with nos. 5 and 2 respectively, which show the corresponding normal curves for a shorter cable. The explanation appears to be that somewhere along the cable the electrical properties changed, causing a region of mismatch; reflection taking place from this region would alter the form of the standing waves at the sending end of the cable and even when the cable was matched at its far end the standing waves would not disappear. The larger variations appearing in no. 3 for a very long cable, and mentioned earlier, were of the same nature and must have been due to faults in the cable.

It is interesting to note here that an alternating-current method of fault location in telephone cables has been described by Palmer and Tufrail (1930) in which the sending-end impedance of a cable is measured at audio frequencies using a bridge circuit. When a fault exists at a single point along the cable, the sending-end impedance shows the same type of variation as in figure 2 (*a*) and the distance of the fault is calculated from the spacing of the maxima. Complications arise when more than one fault is present, as the several reflected waves give rise to such an intricate pattern that the simple calculation can no longer be made. It is, of course, necessary to use audio frequencies for cables of this nature, and on this account the distances involved are greater than at higher frequencies. Roberts (1946) has discussed theoretically both pulse and frequency modulation methods for locating cable faults, with particular reference to high frequencies. He notes that considerable errors may arise using the latter method when more than one fault is present, and concludes that the pulse method can be applied more readily to cable-fault location than a frequency-modulation system giving equal information about the amplitudes and positions of the faults. The paper, which is followed by a comprehensive discussion, may be consulted for a mathematical treatment.

§4. SUMMARY

The frequency-variation method of matching described in this paper has, so far, only been used qualitatively, but it has proved extremely helpful in the adjustment and testing of various types of high-frequency apparatus for which matching is required. The new method is useful in that it will show when and at what frequency and over what range the line is correctly terminated, in addition to which it indicates directly on the screen of an oscilloscope, the relative magnitudes of th standing waves at the sending end of the line over a wide frequency range, and shows visually the effect of adjustments to the impedance terminating the line. Its use in the demonstration of transmission-line properties should be noted, the great advantage here being that, as various adjustments are made to the line and its termination, the instrument gives a continuous visual record on the screen of the oscilloscope. An added advantage is that, in most cases, the results are easily interpreted. The method can be further developed by calibration to enable quantitative results to be obtained; in conjunction with the usual impedance measuring instruments, this would give a fresh approach to certain transmission-line problems.

The application of a similar method to audio frequencies, in connection with fault location in telephone lines, has already been mentioned; at higher frequencies than those employed in the present investigation, where the physical length of line needed would be smaller, it should prove to be as useful as in the applications described in this paper, particularly in the matching of waveguides. The same method can be extended to waves other than electromagnetic waves, e.g. longitudinal sound waves, the requirements being simply a variable frequency source, a "transmission line" and a detector.

ACKNOWLEDGMENTS

I wish to express my gratitude to Dr. N. L. Yates-Fish for his help and advice while the experiments were being performed and for originally suggesting the problem to me.

The work was carried out for the Inter-Services Research Bureau, to whom I am indebted for permission to publish this paper.

REFERENCES

Palmer, W. T., and Tufrail, M. E., 1930. *J. Post Office Elect. Engrs.*, **23**, 42.
Roberts, F. F. 1946. *J. Instn. Elect. Engrs.*, **93**, III, 385.

THERMOELECTRIC POWER OF CADMIUM OXIDE

By J. P. ANDREWS,
Queen Mary College, London

MS. received 28 *January* 1947

ABSTRACT. The thermoelectric e.m.f. (E) of cadmium oxide CdO against platinum is measured over a temperature range of −110° to 800° c. The results above 100° c. are well represented by the formula $E=-aT+b \log_e T-C$, in which the values of the constants a, b, and c vary somewhat according to the previous treatment of the specimen ; their values in a number of different experiments are given. When E is in millivolts, their mean values are $a=0{\cdot}154$, $b=34{\cdot}3$, and $c=225$. The results are discussed, and if it is assumed that the number of conduction electrons is given by $n=Be^{-\beta/kT}$, β is found to be 0·034 electron volts. The sign of the thermoelectric power indicates that CdO is an excess conductor.

§ 1. INTRODUCTION

SEMI-CONDUCTORS, of which CdO is an example, commonly exhibit large thermoelectric (e.m.fs.) against a pure metal ; and if this property is associated with low resistance, as it is in CdO, measurements of thermoelectric power can be made by the ordinary methods applicable to metals. These measurements yield information about the nature of the semi-conductor, and of the current carriers to which the electrical conductivity and thermoelectric properties are due. In terms of the accepted terminology of the subject, this information obtainable may be summarized thus :—

(i) The algebraic sign of the thermoelectric power indicates whether the semi-conductor is to be regarded as of "defect" or "excess" (or alternatively "abnormal" or "normal") type. The conductor will be of normal or excess type if the current is carried mainly by electrons ; and will be of abnormal or defect type if conduction is mainly due to "holes" which behave like positively charged particles. It is worth noting at this point that in a semi-conductor whose current is carried by both electrons and holes, another distinction arises. According to Fowler (1933) a semi-conductor of this kind which is normal at low temperatures retains a negative thermoelectric power at high temperatures ; whereas the thermoelectric power of a semi-conductor which is abnormal at low temperatures will change from a positive value at low temperatures through zero to a negative value at high temperatures. Above a certain ***neutral temperature***, then, the sign of the thermoelectric power would indicate an excess or normal conductor.

(ii) In simple instances, where only one type of carrier need be considered, the thermoelectric power of a semi-conductor against a metal is given (Wilson, 1939) by

$$\frac{dE}{dT} = \mp \frac{\mathbf{k}}{\mathbf{e}} \log \frac{n_2}{n_1} \qquad \ldots\ldots(1)$$

and it is often permissible to ignore the specific contribution of the pure metal In this formula n_2, n_1 are the numbers of electrons or "holes" per unit volume taking part in the process in the semi-conductor and metal respectively, a negative sign being associated with normal or excess conductors.

The equation is of the type derivable from classical theory and is an example of the sort of simplification occurring in semi-conductors, where the number of carriers is so small that the electron gas responsible for the conduction may be considered non-degenerate and having a Maxwell distribution of velocities. It has to be applied to actual semi-conductors with some reserve, since in the first place it implies only one kind of carrier, and in the second place is strictly applicable to metals in which n and n_2 do not vary appreciably with temperature, whereas a fundamental feature of semi-conduction is the increase of n_2 with increase of temperature. However, if the law of intermediate temperatures applies to a thermoelectric circuit containing a semi-conductor as one material, then the change of e.m.f. dE round the circuit when the temperature of the hot junction is raised from T to $T+dT$ (the cold junction being kept at constant temperature) is equal to the e.m.f. in the circuit when the hot junction is at $T+dT$ and the other at T, the mean temperature in this case being very close to T throughout. According to this argument, therefore, the gradient of the e.m.f./temperature curve gives the thermoelectric power which would be obtained from a thermocouple substantially at a temperature T, and the value of n_2 resulting from equation (1) is the electron density in the semi-conductor at a temperature T.

Some of this information is obtainable from other measurements (e.g. of conductivity and Hall effect), which are often preferred to measurements of thermoelectric power because the interpretation offered by current theory appears to be easier, particularly simple relations connecting them with carrier density and charge. It may be, however, that this impression is rather illusory, at any rate where samples in the form of powders are used; for then the shape of the individual grains enters the calculation of both conductivity and Hall effect. This fundamental difficulty is generally ignored, perforce; nevertheless, even in compressed samples, a large uncertainty must remain. Thermoelectric measurements on the other hand are likely to be free from this trouble; and for materials like CdO, the measurements certainly are easier and more accurate than measurements of the Hall effect.

The principal work already carried out on the thermoelectric properties of CdO includes the experiments of Bädeker (1907) on very thin films and over a limited temperature range, and that of Fischer, Dehn and Sustman (1932) for a single temperature difference between hot and cold junctions. Bädeker's experiments yielded a value of 30 to 35 microvolts per degree for the thermoelectric power, while Fischer, Dehn and Sustman (1932) gave 84 microvolts per degree for junction temperatures of 9° and 595° c. Experiments on the allied topic of electrical conduction were carried out by Baumbach and Wagner on bulk material in 1933.

§ 2. MATERIALS

Two varieties of material were used, after the manner of Baumbach and Wagner. These were:—

(*a*) CdO powder, the purest obtainable. A chemical analysis carried out on this powder revealed the following quantities of small impurities:

Lead	0·022 %
Iron	0·009 %
Zinc	0·016 %
Copper	Trace

This material is of a light chocolate brown colour, which darkens temporarily at high temperatures.

(*b*) $CdCO_3$ powder, which, when made into a paste with water, can be moulded into appropriate shapes, dried and calcined, when the brown oxide CdO is obtained. Considerable reduction of volume occurs on heating the white carbonate specimen, and it was in practice impossible to prevent distortion and fissures. The resulting sample of CdO is quite hard and coherent and has a density of about 4·9 gm./c.c., the density of solid CdO being given as 8·1 in the International Critical Tables.

§ 3. METHOD OF MEASUREMENT

Of the methods used, the most fruitful was the following (see figure 1). The specimen in the form of a rod or block, or else of a more or less loose powder contained within a porcelain collar, was placed upright on a thin platinum disc, which in turn was silver-soldered to a brass block on which an electric furnace was wound. The brass block was bored to take the leads and protecting sheath of a chromel-alumel junction which was silver-soldered on to the lower surface of the platinum disc, after passing through a small aperture in the top of the brass block. On top of the specimen, an exactly similar platinum disc was pressed, the disc forming the bottom of a vessel in which cold water circulated. Platinum wires from each disc lead to a potentiometer.

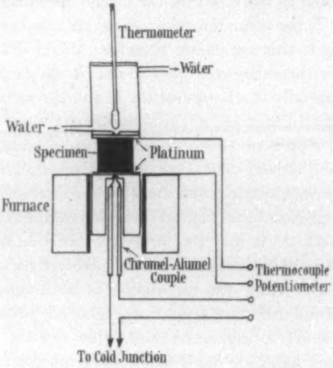


Figure 1.

The chromel-alumel thermo-junction, by which the temperature of the hot end of the specimen was to be measured, was carefully calibrated at a number of fixed points between 0° c. and the melting point of NaCl.

§ 4. RESULTS

In all experiments above room temperature the results resembled those illustrated in the curves of figure 2, but considerable variation occurred between different specimens and for different heat treatments.

In order to summarize the results the curves were fitted to an empirical equation $E = -aT + b.\log_e T - c$, from which

$$\frac{dE}{dT} = -a + \frac{b}{T}.$$

Values of the constants a, b, and c in this equation are given in table 1 for various specimens subjected to different treatments. T is the absolute temperature of the hot end of the specimen.

It may be remarked that the equation

$$E = -aT + b.\log_e T - c$$

used for this summary is not the only type of equation to which the curves may be fitted over a considerable interval of temperature. It is nevertheless true that whether this formula is taken to represent the whole family of curves, or whether it

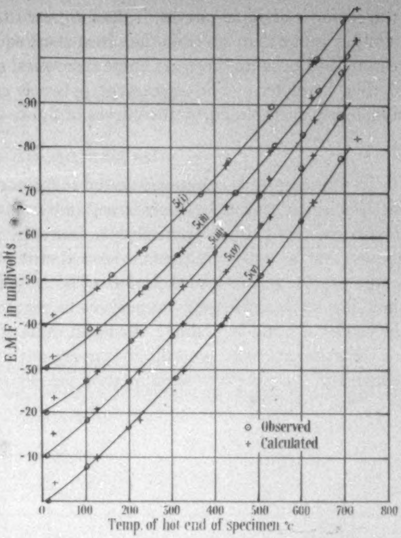


Figure 2. Results of experiment 5 of Table 1. The e.m.f. scale applies to the lowest curve, viz. 5 (v). For the other curves the origin should be raised successively in steps of 10 millivolts.

Table 1

Specimen	Temp. range (°c.)	a (volt/deg. $\times 10^3$)	b (volt $\times 10^3$)	c (volt $\times 10^3$)
1. Loose powder in glass tube	20–460	0·165	34·5	149
2. Stick, from $CdCO_3$, 3 cm. long, heated 4 hours at 800° c.	20–810	0·183	33·7	139
3. Flat pellet from $CdCO_3$	20–700	0·105	26·1	117
4. Same, repeated	20–700	0·114	28·5	128
5. Series of cylinders made from $CdCO_3$:				
(i) heated for 1 hour at 750° c.	20–750	0·155	32·7	140
(ii) ,, ,, 1·75 hours at 750° c.	,,	0·159	34·1	147
(iii) ,, ,, 2·5 ,, ,, ,,	,,	0·185	43·5	198
(iv) ,, ,, 3·25 ,, ,, ,,	,,	0·177	36·0	158
(v) ,, ,, 4·0 ,, ,, ,,	,,	0·170	35·4	150
6. CdO powder packed in a narrow silica tube	20–700	0·153	37·0	166
7. Loose CdO powder in porcelain collar	20–700	0·160	38·3	170
8. Same, repeated	20–700	0·146	31·6	137
9. Pellet, pressed with a hand-press from CdO, after grinding and heating the powder	20–735	0·132	28·9	126
10. Same, repeated	20–720	0·145	34·9	155
11. Machine pressed pellet, from CdO powder. Density 55 % of solid density	20–750	0·160	39·7	176

is used for individual curves, it does appear to fit satisfactorily over a wider temperature range than several other 3-constant equations tried. It is chosen for that reason, and because there are some theoretical grounds for anticipating a relation of this character. The agreement between calculated and observed e.m.f. is sufficiently exemplified by the curves of figure 2, which are quite typical.

§ 5. DISCUSSION OF RESULTS

1. The empirical equation selected suggests that at a temperature $T_m = b/a$, the thermoelectric power becomes zero, and at lower temperatures positive. The interest in this extrapolation derives from the prediction of Fowler, mentioned above, that a semi-conductor whose thermoelectric power is positive at low temperatures, and in which carriers of both signs take part in the effect, ought to show just such a change from positive to negative thermoelectric power. From the values of a and b in table 1, the mean value of T_m should be 225° K., or −48° C.

Experiments were accordingly carried out to continue the curve below room temperature, by the following method (see figure 3). The specimen was firmly held between two platinum discs from which platinum wires led to the potentiometer. The upper disc was maintained at about room temperature by flowing water, while the lower disc was supported on a copper rod capped by a copper disc to bear against the platinum. A Dewar vessel containing liquid air was slowly raised until a chosen length of the rod was immersed. In this way the lower platinum disc was cooled by conduction to a temperature depending on the length of rod left exposed. The temperature was measured as before by a chromel-alumel junction soldered to the platinum disc. The calibration of the couple was checked by measuring the freezing point of absolute alcohol.

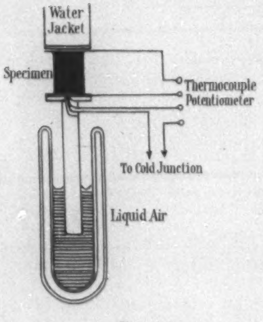


Figure 3.

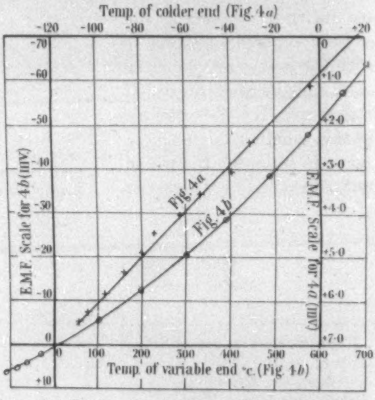


Figure 4. Continuation of the e.m.f. temperature curve to low temperatures.

The result obtained (see figure 4 a) shows that over this restricted range of temperature, the e.m.f. is very nearly proportional to the difference of temperature between the hot and cold ends of the specimen. The complete curve for this specimen, from −110° to 700° c. is given in figure 4 *b*.

Similar results were found for two other specimens.

2. It appears from the foregoing that the equation

$$\frac{dE}{dT} = -a + \frac{b}{T},$$

though applicable over a range of several hundred degrees, cannot be extrapolated to low temperatures. It may be, therefore, that the equation is to be regarded as purely empirical, without much physical significance. On the other hand, however, it is not uncommonly found that the electrons may be excited to the conduction band in a semi-conductor from two different energy levels. The number of conduction electrons may then be supposed given by an expression like

$$n = Ae^{-\alpha/kT} + Be^{-\beta/kT},$$

the activation energies α and β being different. In such a case the term with higher activation energy tends to take over at the higher temperatures, the other at the lower. It is possible that the higher term is dominant in the CdO results above 100° c., while the influence of the lower term begins to appear at lower temperatures (cf. Mott and Gurney, *Electronic Processes in Ionic Crystals*). In the experiments of Baumbach and Wagner (1933) there is indeed some indication of a change in the law of variation of conductivity with temperature, somewhere between 20° and 500° c., and the change is of the kind described. To settle the point for thermoelectric measurements requires experiments over a wider temperature range, but that there is evidence in the present results for a transition of the required type is shown in figure 5 where values of dE/dT are plotted against $1/T$ for three representative experiments. In practically all cases a fairly sharp change in the

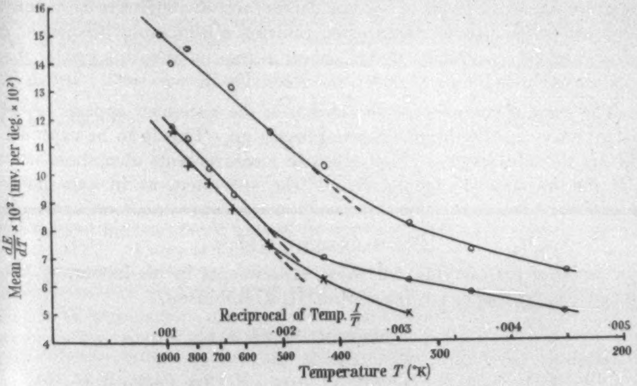


Figure 5. The marked points represent the measured mean values dE/dT over a range fo temperature, usually 50° either side of the temperature T, except at the lowest temperatures where half this range is taken.

slope of such curves appears between 400° and 500° K. In all cases the graphs are approximately rectilinear above that temperature. Assuming then a process like that described above, the value of n_2 in equation (1) may be put equal to $Be^{-\beta/kT}$. It then follows that $\beta = b\,.\,\mathbf{e}$ ergs, where $\mathbf{e}$ is the charge on the electron, or alternatively, $\beta = b$ electron volts. From table 1, the mean value of b then gives the activation energy of β as 0·034 electron volts.

3. Over the temperature range −120° to 800° C. the thermoelectric power is negative against platinum. CdO behaves therefore as an excess or normal semi-conductor.

4. While it is clear that the treatment of the specimen has a considerable effect on the values of the constants *a*, *b*, and *c*—an observation conforming to the usual and often-recorded experience in work on semi-conductors—it is not easy to lay down consistent rules. It is generally true that prolonged heating raises the values, though exceptions appear in table 1. In the series No. 5 of the table, it appears that the constants were increased when the specimen was heated up to a period of 2½ hours, but thereafter decreased. It is generally found that the higher the temperature of heating the more pronounced the effect, and as a rule all three constants increase and decrease together. After prolonged heating at a high temperature, the oxide tends to become darker, and although the darkening fades to some extent on cooling, there often remains a noticeable permanent effect. There is some evidence that a darker colour is associated with a greater stoichiometrical excess of cadmium (Cowley and Rees, 1946).

5. The state of compression or solidity of the specimen appears to have no consistent effect on the thermoelectric properties. This is to be expected, and constitutes the advantage of thermoelectric measurements over those which do depend on the state of aggregation of the specimen, as in experiments on conductivity.

ACKNOWLEDGMENT

For facilities for carrying out these experiments in his laboratory, and for helpful advice, my thanks are due to Prof. H. R. Robinson.

REFERENCES

Bädeker, K., 1907. *Ann. Phys., Lpz.*, **22,** 749.
von Baumbach, H. H. and Wagner, C., 1933. *Z. Phys. Chem.*, B, **22,** 199.
Cowley, J. M. and Rees, A. L. G., 1946. *Nature, Lond.*, **158,** 550.
Fischer, F., Dehn, K. and Sustmann, H., 1932. *Ann. Phys., Lpz.*, **15,** 109.
Fowler, R. H., 1933. *Proc. Roy. Soc.*, A, **140,** 505.
Wilson, A. H., 1939. *Metals and Semi-Conductors*, p. 56.

DISCUSSION

Prof. G. I. Finch. I wish to draw attention to the possible effects of crystal structure and particularly crystal habit variations on the results. Cadmium oxide is of rock-salt structure and it is quite probable that compression, heating and so forth might well change the crystal habit in a micro-crystalline mass and hence affect the values of Dr. Andrews' constants.

Dr. H. H. Hopkins. Perhaps we could be told whether Dr. Andrews is quite satisfied that the variations in the values of the constants of his equation do not derive from his method of " fitting " the equation to the experimental results, and that they do, in fact, indicate different physical conditions in the various samples under test ?

Mr. J. H. **Awbery**. In fitting equations to his results, Dr. Andrews starts with the equation

$$\frac{dE}{dT} \times a + \frac{b}{T}, \qquad \ldots\ldots \text{(i)}$$

of which the integrated form is

$$E = aT + b\,.\,\ln T + c, \qquad \ldots\ldots \text{(ii)}$$

where E is the thermal e.m.f. when the hot side of the specimen is at temperature T, the cold side remaining always at a temperature T_0.

In presenting the paper, he showed a slide of E versus T, with points calculated from the equation marked on it, and pointed out that in every case, the calculated point for the lowest temperature lay above the experimental curve. This is a serious matter, for it means that if the curve calculated from his equation were plotted, it would not give $E=0$ at $T=T_0$. This is clear, because according to (i) dE/dT is monotonic, and thus the curve representing (ii) cannot have a hump on it, which would correspond to a minimum of dE/dT. Without the hump, it cannot be correct at higher temperatures, high at lower ones, and correct at T_0. It follows that Dr. Andrews' calculated curve cannot be correct, and further, that the constants in the equation (ii) to this curve cannot be correct.

The error could arise if the thermometer with which he measured the temperature of the cold face did not agree with the thermocouple with which he measured that of the hot face, or if either did not in fact read the temperature of the face of the specimen, but a more likely explanation (especially in view of the author's skill as an experimenter) is that the calculated curve does not in fact properly represent the observations.

It was mentioned in the course of the discussion on the paper that the "calculated" curve was obtained by fitting an equation of the form (ii) by selecting three points (presumably on a curve drawn graphically). I would suggest that a better procedure, whenever there are conditions which a curve must satisfy, is to enforce compliance with these conditions first. Thus equation (ii) requires

$$0 = aT_0 + b\,.\,\ln T_0 + c,$$

whence by subtraction

$$E = a(T - T_0) + b\,.\,\ln T/T_0 \qquad \ldots\ldots \text{(iii)}$$

or

$$\frac{E}{T - T_0} \times a + \frac{b\,.\,\ln T/T_0}{T - T_0} \qquad \ldots\ldots \text{(iii } a\text{)}$$

In this form, the equation should give a linear plot of $E/(T-T_0)$ against $(\ln T/T_0)/(T-T_0)$, and the slope of the line should give b, its intercept giving a. c would then follow as $-(aT_0 + b\,.\,\ln T_0)$. I suggest that by this procedure, all points could be given due weight, that the curve would necessarily comply with the condition at T_0, and that the values of a, b and c deduced would thus be more reliable.

Prof. R. W. **Ditchburn**. Do I correctly understand (*a*) that the curves were fitted by the method of least squares and (*b*) that the differences between samples are large compared with the difference between measurements on the same sample?

Author's reply. I think it very likely that Prof. Finch is right, and that some part of the variation found ought to be ascribed to the processes he mentioned. The same is true of nearly all measurements on semi-conductors—and, I suppose, on solids in general.

I would like to refer Mr. Awbery to the paper itself, when a printed copy reaches him; for I believe he will find his criticisms already dealt with there. In fact, the matter of the absence of a minimum in dE/dT, which he feels is so serious, is regarded in the same light in my paper, and was made a prominent part of my verbal account to the Society; and I am grateful to him for adding such skilful re-emphasis. In the paper he will see that recognition of the failure of this test of the empirical equation led to further experiments and a suggested revision or restriction of the equation. Perhaps he would agree that under the circumstances, provided the equation is fitted as well as the curves show that it was, the particular method of arriving at the values of the constants is of secondary importance. That is not to say that his suggested method is ungratefully received, or will not be made good use of.

As Dr. Hopkins suggests, I daresay it is possible that with a different method of computation of the constants of the equation, there might be some slight difference in the variations I have given; but the effects are large enough to leave no doubt of their existence.

I would not, however, be inclined to base any important conclusions on the numerical values given for these changes until more is understood about their origin, or about the right kind of experimental condition for their determination.

In reply to Prof. Ditchburn I may say that the method of least squares was not used. It is not easy to give a short answer to the second question, but I believe Prof. Ditchburn will be able to judge the relative differences on consulting the annotated table of results given in the paper.

THE CALCULATION OF POTENTIAL-ENERGY CURVES FROM BAND-SPECTROSCOPIC DATA

By A. L. G. REES,
Division of Industrial Chemistry, Council for Scientific and
Industrial Research, Melbourne, Australia

MS. received 15 *January* 1947

ABSTRACT. Analytical expressions equivalent to the Rydberg-Klein graphical construction have been derived for the calculation of potential-energy curves of diatomic molecules from band-spectroscopic data. These expressions have been shown to lead to satisfactory values for D_e and r_e and, in contrast to the Rydberg method, enables the potential-energy (p.e.) curve to be evaluated with accuracy in the region of the minimum. The Morse function has been shown to satisfy the analytic expressions where a quadratic in $(v+\frac{1}{2})$ has been used to express the band data. A method for dealing with electronic states in which a discontinuity in the law of force occurs has been outlined. By using a series of quadratics to satisfy the band data, the calculation of p.e. curves by the use of expressions given in this paper is reduced to a rapid and accurate procedure. The methods outlined have been illustrated by the computation by several procedures of the p.e. curve for the $^3\Pi_{0^+u}$ state of the bromine molecule.

§1. INTRODUCTION

Several closed expressions have been suggested to facilitate the construction of a potential-energy curve for a particular electronic state of a diatomic molecule (Morse, 1929; Rydberg, 1931; Rosen and Morse, 1932); these have been chosen to lead to an exact solution of the Schrödinger equation (Morse, 1929; Rosen and Morse, 1932) or have been derived as the result of approximation (Rydberg, 1931). For the interpretation of much of the band-spectroscopic data it is necessary to construct more accurate functions than those referred to, and recourse is then had to the laborious graphical procedure of Oldenberg (1929), Rydberg (1931 and 1933) and Klein (1932). This method cannot be applied with accuracy at low vibrational quantum numbers, for which the power-series expansion method of Crawford and Jorgensen (1936) must be employed. The power series for the potential energy in terms of the displacement of the atoms from their equilibrium position does not converge rapidly for any range of displacements, and involves even more computation than the Rydberg-Klein method. Taking advantage of the fact that a cubic in the vibrational quantum number will describe with sufficient accuracy the spectroscopic data for vibrational states, an analytical expression has been derived which is equivalent to the Rydberg-Klein procedure and which facilitates the construction of p.e. curves considerably. In this paper the Rydberg method has been reduced to an analytical operation and the procedure for dealing with cases in which a change in the law of force occurs at high vibrational quantum numbers is outlined.

§2. DERIVATION OF EXPRESSIONS RELATING POTENTIAL ENERGY AND INTERNUCLEAR SEPARATION

The Rydberg-Klein method leads to an expression (Klein, 1932)

$$r_{1,2}(U)=(f/g+f^2)^{1/2}\pm f \qquad \ldots\ldots(1)$$

for the maximum and minimum values $r_{1,2}$ of the interatomic distance for a molecule vibrating with an energy U. f and g are defined in terms of the function

$$S(U,\kappa)=\frac{1}{\pi(2\mu)^{1/2}}\int_0^{I'}\{U-E(I,\kappa)\}^{1/2}\,dI, \qquad \ldots\ldots(2)$$

in which $E(I,\kappa)$ is the sum of the vibrational and rotational energy of the molecule,

$$I=\boldsymbol{h}(v+\tfrac{1}{2})^*,$$

$$\kappa=\frac{J(J+1)\boldsymbol{h}^2}{8\pi^2\mu}=\frac{J(J+1)}{b}{}^*,$$

v the vibrational quantum number,
J the rotational quantum number,
μ the reduced mass of the molecule,
and $I=I'$ when $U=E$.

f and g are then defined by

$$f=\frac{\partial S}{\partial U},$$

$$g=-\frac{\partial S}{\partial \kappa}.$$

Since $E(I,\kappa)$ can be expressed in terms of v and J and the derived constants ω_e, $\omega_e x_e$, $\omega_e y_e$, B_e, α and D_e to the accuracy of the experimental data, in some cases by a quadratic in v only, but in most by a cubic, then the expression to be integrated is known. If the integration can be performed, the function $r(U)$ can be obtained in terms of the band-spectroscopic constants.

$E(I,\kappa)$ *a quadratic in I*

$E(I,\kappa)$ may be expressed approximately as

$$\omega_e(v+\tfrac{1}{2})-\omega_e x_e(v+\tfrac{1}{2})^2+B_e J(J+1)+D_e J^2(J+1)^2-\alpha J(J+1)(v+\tfrac{1}{2}). \qquad \ldots\ldots(3)$$

Substituting this into (2) and introducing the variables I and κ, we have

$$S(U,\kappa)=q\int_0^{I'}(A-lI+mI^2)^{1/2}\,dI,$$

where

$$1/q=\pi(2\mu\boldsymbol{h})^{1/2},$$

$$A=\boldsymbol{h}\{U-B_e J(J+1)-D_e J^2(J+1)^2\},$$

$$l=\omega_e-\alpha J(J+1),$$

and

$$m=\frac{\omega_e x_e}{\boldsymbol{h}}.$$

---

* The expressions given here are the quantum-mechanical equivalents of the classical quantities I and κ.

Differentiating with respect to U and κ under the integral sign and then performing the integrations leads to the following expressions for f and g:

$$f=\frac{q\boldsymbol{h}}{2m^{1/2}}\log_e\left\{\frac{(l^2-4mA)^{1/2}}{l-(4mA)^{1/2}}\right\},$$

$$g=\frac{qb}{4m^{3/2}}\left[\alpha(4mA)^{1/2}+(2mB_e\boldsymbol{h}+4mD_e\boldsymbol{h}b\kappa-\alpha l)\log_e\left\{\frac{(l^2-4mA)^{1/2}}{l-(4mA)^{1/2}}\right\}\right],$$

which for U and the constants ω_e, $\omega_e x_e$, α, B_e and D_e expressed in wave numbers and for the rotationless state ($J=0$) become

$$f(\text{cm.})=\left(\frac{\boldsymbol{h}}{8\pi^2\mathbf{c}\mu\,.\,\omega_e x_e}\right)^{1/2}\log_e\left\{\frac{(\omega_e{}^2-4\omega_e x_e U)^{1/2}}{\omega_e-(4\omega_e x_e U)^{1/2}}\right\}, \qquad \ldots\ldots(4)$$

$$g(\text{cm.}^{-1})=\left(\frac{2\pi^2\mu\mathbf{c}}{\boldsymbol{h}(\omega_e x_e)^3}\right)^{1/2}\Bigg[\alpha(4\omega_e x_e U)^{1/2}$$

$$+(2\omega_e x_e B_e-\alpha\omega_e)\log_e\left\{\frac{(\omega_e{}^2-4\omega_e x_e U)^{1/2}}{\omega_e-(4\omega_e x_e U)^{1/2}}\right\}\Bigg]. \qquad \ldots\ldots(5)$$

$E(I,\kappa)$ a cubic in I

If the law of force remains unchanged for all vibrational levels, then an expression of the form

$$\omega_e(v+\tfrac{1}{2})-\omega_e x_e(v+\tfrac{1}{2})^2+\omega_e y_e(v+\tfrac{1}{2})^3+B_eJ(J+1)$$
$$+D_eJ^2(J+1)^2-\alpha J(J+1)(v+\tfrac{1}{2}) \qquad \ldots\ldots(6)$$

will describe the energies of all the vibrational-rotational states adequately. The function $S(U,\kappa)$ then becomes

$$q\int_0^{I'}(A-lI+mI^2+nI^3)^{1/2}\,dI,$$

where

$$n=-\frac{\omega_e y_e}{\boldsymbol{h}^2}$$

and A, l, m and q are defined as before. It is convenient to differentiate $S(U,\kappa)$ with respect to U and κ before integrating with respect to I. We then obtain

$$f=\frac{\partial S}{\partial U}=\frac{q\boldsymbol{h}}{2}\int_0^{I'}\frac{dI}{(A-lI+mI^2+nI^3)^{1/2}} \qquad \ldots\ldots(7)$$

and

$$g=-\frac{\partial S}{\partial\kappa}=\frac{q(B_eb\boldsymbol{h}+2D_eb^2\boldsymbol{h}\kappa)}{2}\int_0^{I'}\frac{dI}{(A-lI+mI^2+nI^3)^{1/2}}$$
$$-\frac{\alpha bq}{2}\int_0^{I'}\frac{I\,.\,dI}{(A-lI+mI^2+nI^3)^{1/2}} \qquad \ldots\ldots(8)$$

(1) *Expressions for f*

By making the substitution $z=I+m/3n$ we obtain

$$f=\frac{q\boldsymbol{h}}{n^{1/2}}\int_{m/3n}^{I'+m/3n}\frac{dz}{(4z^3-g_2z-g_3)^{1/2}},$$

where

$$g_2=4\left(\frac{l}{n}+\frac{m^2}{3n^2}\right)$$

and

$$g_3=-4\left(\frac{lm}{3n^2}+\frac{2}{27}\frac{m^3}{n^3}+\frac{A}{n}\right)$$

are the invariants of the cubic.

Making the substitution $z=\wp(u\,;\,g_2,g_3)$, where $\wp(u\,;\,g_2,g_3)$ is the Weierstrassian elliptic function,

$$f=\frac{qh}{n^{1/2}}\int_{u_1}^{u_2}\frac{\wp'(u)\,du}{(4\wp^3(u)-g_2\wp(u)-g_3)^{1/2}}$$
$$=\frac{qh}{n^{1/2}}(u_2-u_1),$$

since $\wp'^2(u)=4\wp^3(u)-g_2\wp(u)-g_3$.

The form of the solution is determined by (i) the sign of the discriminant $\Delta(=g_2^3-27g_3^2)$ of the cubic, and (ii) the sign of $\omega_e y_e$.

A real dissociation energy is possible only for the two cases in which $\Delta>0$; however, where discontinuities occur in the law of force, ranges of v-values for which $\Delta<0$ may be encountered; the solutions for these conditions will therefore be indicated. The solutions will be given in detail for $\omega_e y_e$ negative only; those for $\omega_e y_e$ positive are given in table 1.

(*a*) *For $\omega_e y_e$ negative*; $\Delta>0$,

$$\wp(u)=e_3+(e_1-e_3)\,\mathrm{ns}^2\,(u(e_1-e_3)^{1/2}\,|\,k),$$

where
$$k=\left(\frac{e_2-e_3}{e_1-e_3}\right)^{1/2}$$

is the modulus of the Jacobian elliptic function ns, and e_1, e_2 and e_3 are the real roots of the cubic, such that $e_1>e_2>e_3$ and $e_1+e_2+e_3=0$.

When $I=I'$, $z=e_2=\wp(u)$. Since $e_2\to m/3n$ as $U\to 0$, e_2 is the physically significant root.

Hence
$$u_2=(e_1-e_3)^{-1/2}\,\mathrm{sn}^{-1}\,(1/k\,|\,k),$$

and since k is always <1, u_2 is unreal and may be written

$$u_2=(e_1-e_3)^{-1/2}(K+iK').$$

K and K' are the complete elliptic integrals defined by the hypergeometric functions

$$\frac{1}{2\pi}F(\tfrac{1}{2},\tfrac{1}{2}\,;\,1\,;\,k^2)\quad\text{and}\quad\frac{1}{2\pi}F(\tfrac{1}{2},\tfrac{1}{2}\,;\,1\,;\,k'^2)$$

respectively $(k'^2=1-k^2)$.

When $I=0$, $z=m/3n=\wp(u)$, so that

$$u_1=(e_1-e_3)^{-1/2}\,\mathrm{sn}^{-1}\left\{\left(\frac{e_1-e_3}{m/3n-e_3}\right)^{1/2}\middle|\,k\right\},$$

and since $e_3<0$ and $e_1>m/3n$, u_1 is complex also.

However, using the relation (Whittaker and Watson, 1940 c)

$$\mathrm{sn}\,(v+iK'\,|\,k)=k^{-1}\,\mathrm{ns}\,(v\,|\,k),$$

we obtain
$$u_1=(e_1-e_3)^{-1/2}\left[\mathrm{sn}^{-1}\left\{\left(\frac{m/3n-e_3}{e_2-e_3}\right)^{1/2}\middle|\,k\right\}+iK'\right].$$

Expressing U and the spectroscopic constants in wave numbers, we obtain

$$f=q\left(\frac{h}{-\omega_e y_e\,.\,\mathbf{c}}\right)^{1/2}(\theta_1-\phi_1),\qquad\ldots\ldots(9)$$
$$\theta_1=(e_1-e_3)^{-1/2}K,$$
$$\phi_1=(e_1-e_3)^{-1/2}\,\mathrm{sn}^{-1}\left\{\left(\frac{m/3n-e_3}{e_2-e_3}\right)^{1/2}\middle|\,k\right\},$$

which is real.

Table 1

$$f=q\left(\frac{\boldsymbol{h}}{-\omega_e y_e\,.\,\mathbf{c}}\right)^{1/2}(\theta-\phi),$$

$$g=4\pi\left(\frac{2\mu\boldsymbol{c}}{-\omega_e y_e}\right)^{1/2}\left[\left(B_e-\frac{\alpha\,.\,\omega_e x_e}{3\omega_e y_e}\right)(\theta-\phi)+\frac{\alpha}{\boldsymbol{h}}(\zeta(\theta)-\zeta(\phi)+\psi)\right]$$

$\triangle$	y_e	Real roots	Physically significant root	θ	ϕ	ψ
>0	<0	$e_1>e_2>e_3$	$e_2(>0)$	$K(e_1-e_3)^{-1/2}$	$(e_1-e_3)^{-1/2}\mathrm{sn}^{-1}\left\{\left(\frac{m/3n-e_3}{e_1-e_3}\right)^{\frac{1}{2}}\middle\vert k\right\}$	$\left\{\frac{(e_2-m/3n)(e_1-m/3n)}{(m/3n-e_3)}\right\}^{1/2}$
<0	<0	e_2	$e_2(>0)$	$KH^{-1/2}$	$\frac{1}{2}H^{-1/2}\mathrm{cn}^{-1}\left\{-\left(\frac{H-m/3n+e_2}{H+m/3n-e_2}\right)\middle\vert k\right\}$	0
>0	>0	$e_1>e_2>e_3$	$e_3(<0)$	$iK'(e_1-e_3)^{-1/2}$	$i(e_1-e_3)^{-1/2}\mathrm{sn}^{-1}\left\{\left(\frac{e_1-e_3}{e_1-m/3n}\right)^{\frac{1}{2}}\middle\vert k'\right\}^{*}$	0
<0	>0	e_2	$e_2(<0)$	$iK'H^{-1/2}$	$\frac{i}{2}H^{-1/2}\mathrm{cn}^{-1}\left\{-\left(\frac{H+m/3n-e_2}{H-m/3n+e_2}\right)\middle\vert k'\right\}^{*}$	0

* Where θ and ϕ are imaginary quantities, both f and g are real, since $\zeta(iu)=i\,.\,\zeta(u)$.

(*b*) *For $\omega_e y_e$ negative*; $\Delta<0$, there is only one real root, designated e_2, and

$$\wp(u)=e_2+H\frac{1+\operatorname{cn}(2uH^{1/2}|k)}{1-\operatorname{cn}(2uH^{1/2}|k)},$$

where
$$H^2=2e_2{}^2+g_3/4e_2$$
and
$$k^2=\tfrac{1}{2}-\tfrac{3}{4}\,.\,e_2/H,$$
from which we obtain (U, ω_e, etc., in cm.$^{-1}$),

$$f=q\left(\frac{\boldsymbol{h}}{-\omega_e y_e\,.\,\mathbf{c}}\right)^{1/2}(\theta_2-\phi_2), \qquad \ldots\ldots(10)$$

$$\theta_2=H^{-1/2}K,$$

$$\phi_2=\tfrac{1}{2}H^{-1/2}\operatorname{cn}^{-1}\left\{-\left(\frac{H-m/3n+e_2}{H+m/3n-e_2}\right)\Big|\,k\right\}.$$

(2) *Expressions for g*

Inspection of equation (8) shows that the integral involved in the first term of g is identical with that for f, and it may be written, for $\Delta>0$ and $\omega_e y_e$ negative, as

$$\frac{qb\boldsymbol{h}(B_e-2D_e b\kappa)}{n^{1/2}}(\theta_1-\phi_1).$$

On making the substitution $I+m/3n=z=\wp(u;\ g_2, g_3)$, the second term becomes

$$\frac{\alpha bq}{n^{1/2}}\cdot\frac{m}{3n}(\theta_1-\phi_1)-\frac{\alpha bq}{n^{1/2}}\int_{u_1}^{u_2}\wp(u)\,.\,du;$$

$$\int_{u_1}^{u_2}\wp(u)\,.\,du=\zeta(u_2)-\zeta(u_1),$$

where $\zeta(u)$ is the Weierstrassian ζ-function.

Although u_1 and u_2 are complex variables, by making use of the relations

$$\zeta(u+v)=\zeta(u)+\zeta(v)+\tfrac{1}{2}\left\{\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)}\right\}$$

and
$$\wp(u;\ g_2, g_3)=\lambda\,.\,\wp(u\lambda^{1/2};\ g_2\lambda^{-2},\ g_3\lambda^{-3})$$

(Whittaker and Watson, 1940 a and b), the pure imaginary parts of the functions may be separated and cancelled, leading to

$$\zeta(u_1)-\zeta(u_2)=\zeta(\theta_1)-\zeta(\phi_1)+\psi;\quad\left\{\frac{(e_2-m/3n)(e_1-m/3n)}{(m/3n-e_3)}\right\}^{1/2}=\psi,$$

which is real.

For the rotationless state ($\kappa=0$), and with U, ω_e, etc. in cm.$^{-1}$, we have

$$g=4\pi\left(\frac{2\mu\mathbf{c}}{-\omega_e y_e}\right)^{1/2}\left[\left(B_e-\frac{\alpha\omega_e x_e}{3\omega_e y_e}\right)(\theta_1-\phi_1)+\frac{\alpha}{\boldsymbol{h}}(\zeta(\theta_1)-\zeta(\phi_1)+\psi)\right] \qquad \ldots\ldots(11)$$

For a negative discriminant ($\Delta<0$), and $\kappa=0$,

$$g=4\pi\left(\frac{2\mu\mathbf{c}}{-\omega_e y_e}\right)^{1/2}\left[\left(B_e-\frac{\alpha\omega_e x_e}{3\omega_e y_e}\right)(\theta_2-\phi_2)+\frac{\alpha}{\boldsymbol{h}}(\zeta(\theta_2)-\zeta(\phi_2))\right] \qquad \ldots\ldots(12)$$

The solutions for $\omega_e y_e$ positive are given in table 1.

§3. THE DISSOCIATION ENERGY

The condition for $U=\boldsymbol{D}_e$*, the dissociation energy, is that f, the half-width of the $U(r)$ curve at an energy U, should become infinite. For the quadratic case, this can be seen to occur when $\omega_e^2=4\omega_e x_e \, . \, U$ in equation (4), i.e.

$$\boldsymbol{D}_e=\frac{\omega_e^2}{4\omega_e x_e},$$

as expected.

For the cubic case, only $\Delta>0$ can lead to a real dissociation energy. In equation (9), $f\to\infty$ as $K\to\infty$, which occurs when $k\to 1$. Since

$$k=\left(\frac{e_2-e_3}{e_1-e_3}\right)^{1/2},$$

$k=1$ only when $e_1=e_2$, i.e. when $\Delta=0$ or $27g_3^2=g_2^3$.

We then have

$$U=\boldsymbol{D}_e=\frac{1}{27\omega_e y_e^2}\{2(\omega_e x_e^2-3\omega_e\,.\,\omega_e y_e)^{3/2}-\omega_e x_e(2\omega_e x_e^2-9\omega_e\,.\,\omega_e y_e)\}, \qquad \ldots\ldots(13)$$

which is consistent with the value obtained by equating $\partial E/\partial v$ to zero to obtain $v_{\max}$ and substituting to obtain $E_{v_{\max}}=\boldsymbol{D}_e$.

§4. THE EQUILIBRIUM INTERNUCLEAR SEPARATION

From equation (1) it is evident that the equilibrium internuclear distance r_e corresponds to the limit

$$\underset{U\to 0}{\mathrm{Lt}}\{(f/g+f^2)^{1/2}\pm f\}=\frac{1}{\underset{U\to 0}{\mathrm{Lt}}\, g/f}.$$

Equations (4) and (5) are readily shown to lead to

$$r_e=\left(\frac{\boldsymbol{h}}{8\pi^2\mathbf{c}B_e\mu}\right)^{1/2},$$

in agreement with definition.

For the cubic case, equations (9) and (11) lead to

$$\underset{U\to 0}{\mathrm{Lt}}\, g/f=b\left(B_e+\frac{\alpha m}{3n}\right)-\frac{\alpha b}{\boldsymbol{h}}\,.\,e_1+\frac{\alpha b}{\boldsymbol{h}}(e_1-m/3n)$$

$$=b\,.\,B_e.$$

As $b=\dfrac{8\pi^2\mu}{\boldsymbol{h}^2}$, $r_e=\left(\dfrac{\boldsymbol{h}}{8\pi^2\mathbf{c}B_e\mu}\right)^{1/2}$, as before.

§5. THE MORSE FUNCTION

The function proposed by Morse may be shown to lead to the same expression for f as equation (4). The Morse function may be written as

$$(r_{1,2}-r_e)=-\frac{1}{a}\log_e\left\{1\pm\left(\frac{U}{\boldsymbol{D}_e}\right)^{1/2}\right\},$$

where

$$a=2\pi\left(\frac{2\mu\,.\,\omega_e x_e\,.\,\mathbf{c}}{\boldsymbol{h}}\right)^{1/2}.$$

---

* $\boldsymbol{D}_e$, the dissociation energy, should not be confused with D_e, one of the coefficients in the rotational term.

The width of this curve at a vibrational level of potential energy U is

$$2f=(r_2-r_e)-(r_1-r_e)$$
$$=\frac{1}{a}\log_e\left\{\frac{\boldsymbol{D}_e-U}{(\boldsymbol{D}_e^{1/2}-U^{1/2})^2}\right\},$$

whence
$$f=\left(\frac{\boldsymbol{h}}{8\pi^2\mathbf{c}\mu\omega_e x_e}\right)^{1/2}\log_e\left\{\frac{(\boldsymbol{D}_e-U)^{1/2}}{\boldsymbol{D}_e^{1/2}-U^{1/2}}\right\},$$

which is identical with equation (4). This equivalence is to be expected, since the Morse function leads to an exact solution of the one-dimensional Schrödinger equation and describes the energy levels E_v as a quadratic in $(v+\frac{1}{2})$, whereas in the derivation of equation (4) this has been used as the fundamental assumption.

§ 6. DISCONTINUITY IN THE LAW OF FORCE

Birge (1929 b) has pointed out that a discontinuity in the law of force for a certain molecular vibrational energy is of frequent occurrence and, on either side of this discontinuity, the use of functions having different values of ω_e, $\omega_e x_e$, etc. is necessitated if the band spectroscopic data are to be satisfied. In the general case, then, we have

$$0<I<I_1; \qquad A_1, l_1, m_1, n_1;$$
$$I_1<I<I_2; \qquad A_2, l_2, m_2, n_2.$$

. . . .

Writing $(A_i-l_iI+m_iI^2+n_iI^3)^{-1/2}=\chi_i$,

$$f=\frac{q\boldsymbol{h}}{2}\left\{\int_{I_r}^{I'}\chi_{r+1}\,.\,dI+\sum_{i=1}^{i=r}\int_{I_{i-1}}^{I_i}\chi_i\,.\,dI\right\}, \qquad \ldots\ldots(15)$$

$$g=\frac{q(B_e b\boldsymbol{h}+2D_e b^2\boldsymbol{h}\kappa)}{2}\left\{\int_{I_r}^{I'}\chi_{r+1}\,.\,dI+\sum_{i=1}^{i=r}\int_{I_{i-1}}^{I_i}\chi_i\,.\,dI\right\}$$
$$-\frac{q\alpha b}{2}\left\{\int_{I_r}^{I'}\chi_{r+1}\,.\,I\,.\,dI+\sum_{i=1}^{i=r}\int_{I_{i-1}}^{I_i}\chi_i\,.\,I\,.\,dI\right\} \qquad \ldots\ldots(16)$$

in which $I_0=0$.

§ 7. THE CONSTRUCTION OF P.E. CURVES

It is apparent that the potential function $U(r)$ can now be constructed from the spectroscopic constants by means of the expressions (9) and (11), (10) and (12) or (15) and (16) as required. The Jacobian elliptic functions sn and cn and the elliptic integral K have been tabulated (Milne-Thomson, 1931) and a double inverse interpolation of these tables is involved in the computation; the Weierstrassian ζ-function, however, has been tabulated for the equianharmonic case only and it must therefore be evaluated by means of the series

$$\zeta(u)=\frac{1}{u}-\frac{g_2u^3}{60}-\frac{g_3u^5}{140}-\frac{g_2^{\,2}u^7}{8400}-\frac{g_2g_3}{18480}u^9$$
$$-\left(\frac{g_2^{\,3}}{1{\cdot}716\times 10^6}+\frac{g_3^{\,2}}{1{\cdot}12112\times 10^5}\right)u^{11}-\frac{g_2^{\,2}g_3}{2{\cdot}4024\times 10^6}u^{13}-\ldots.$$
$$\ldots\ldots(17)$$

Normally, evaluation to the term in u^{13} is sufficient. This computation can be performed more rapidly and accurately than the Rydberg graphical procedure, although the evaluation of the ζ-function is laborious.

A rapid method, with somewhat less theoretical justification, but with adequate accuracy, may be used to avoid the evaluation of the ζ-function. Since $\omega_e y_e$ is invariably much smaller than $\omega_e x_e$, it is possible to represent the band data by a small number of quadratic expressions in $(v+\frac{1}{2})$ satisfactorily. The method is then analogous to that outlined in §6, but some of the points of discontinuity in the apparent ω_e and $\omega_e x_e$ values will be artificial and will not necessarily indicate a change in the law of force. After fixing the points of discontinuity by inspection of a plot of $(G(v))/(v+\frac{1}{2})$ against $(v+\frac{1}{2})$, the constants ω_e and $\omega_e x_e$ for each continuous range may be evaluated by the least squares method. The subsequent calculation of $r_{1,2}(U)$ is extremely rapid.

§8. ILLUSTRATIVE CALCULATION

Using Brown's (1931) data for the ${}^3\Pi_{0^+u}$ state of ${}^{79}\mathrm{Br}\,{}^{81}\mathrm{Br}$, the potential-energy curve for the rotationless state ($J=0$) has been evaluated up to the level $v'=21$ by several different procedures for comparison. The procedures employed were:—

(i) Using the values $\omega_e=165{\cdot}4\,\mathrm{cm}^{-1}$, $\omega_e x_e=1{\cdot}577\,\mathrm{cm}^{-1}$, $\omega_e y_e=-0{\cdot}0087\,\mathrm{cm}^{-1}$, $B_e=0{\cdot}0595\,\mathrm{cm}^{-1}$ and $\alpha=6{\cdot}25\times10^{-4}\,\mathrm{cm}^{-1}$ in equation (6) to express the spectroscopic data adequately over the range $v'=0$ to $v'=21$, the quantities f and g were evaluated by means of equations (9) and (11). r_1 and r_2 values were then calculated by means of equation (1) for each $U=E_{v'}$.

(ii) The values $\omega_e=165{\cdot}4\,\mathrm{cm}^{-1}$, $\omega_e x_e=1{\cdot}59\,\mathrm{cm}^{-1}$, $r_e=2{\cdot}656\times10^{-8}\,\mathrm{cm}$. and the extrapolated value of $\boldsymbol{D}_e=3814\,\mathrm{cm}^{-1}$ were used to construct a Morse curve.

(iii) Using $\omega_e=165{\cdot}4\,\mathrm{cm}^{-1}$ and $\omega_e x_e=1{\cdot}59\,\mathrm{cm}^{-1}$, the potential-energy curve was computed from equations (4) and (5). This procedure is equivalent to constructing a Morse curve in which

$$\boldsymbol{D}_e=\frac{\omega_e^{\,2}}{4\omega_e x_e}.$$

(iv) Using values of ω_e, $\omega_e x_e$ and derived $\boldsymbol{D}_e$ and r_e values appropriate to each pair of levels $E_{v'}$ and $E_{v'+1}$, separate Morse functions have been evaluated for each region $E_{v'}<U<E_{v'+1}$, so that the complete potential-energy curve is built up from 21 segments of separate Morse curves. This is equivalent to assuming an artificial discontinuity in the law of force at each level and calculating according to equations (15) and (16) in which $\chi_i=(A_i-l_i I+m_i I^2)^{-1/2}$.

The results obtained by application of each of these methods are tabulated for $U=E_{v'}$ up to $v'=21$ in table 2 and plotted in the figure for comparison.

It is evident that method (i) and (iv) give results within the accuracy expected from method (i), which is limited by the incomplete evaluation of the Weierstrassian ζ-function (taken to the term in u^{17}), and that both the Morse functions (ii) and (iii) diverge markedly from the correct function, particularly on the outer limit and at the higher vibrational levels.

Brown (1931) found a discontinuity in the law of force to occur at $v'=21$; examination of his data indicates that a second discontinuity occurs at $v'=37$.

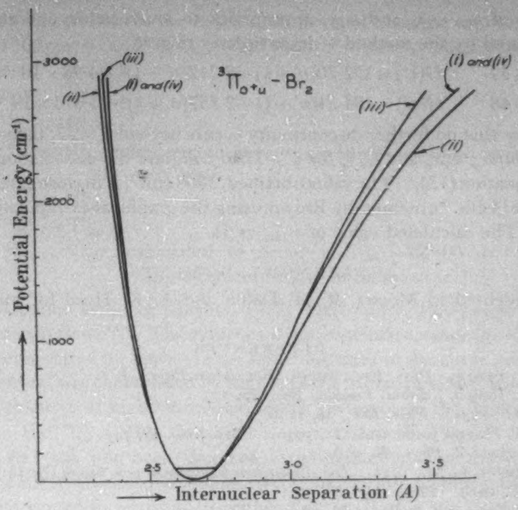


Table 2

v'	U (cm.$^{-1}$)	r_1 (A.)				r_2 (A.)			
		(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
0	82·3	2·589	2·586	2·589	2·590	2·731	2·738	2·733	2·732
1	244·5	2·548	2·547	2·549	2·550	2·797	2·805	2·796	2·796
2	403·5	2·516	2·511	2·519	2·519	2·844	2·858	2·844	2·845
3	559·2	2·495	2·489	2·498	2·499	2·884	2·903	2·888	2·887
4	711·6	2·478	2·471	2·481	2·482	2·922	2·947	2·924	2·925
5	860·5	2·464	2·456	2·467	2·467	2·959	2·975	2·960	2·962
6	1006·1	2·452	2·443	2·454	2·455	2·997	3·027	2·997	3·000
7	1148·1	2·442	2·431	2·442	2·444	3·028	3·066	3·030	3·034
8	1286·7	2·433	2·420	2·432	2·435	3·065	3·104	3·063	3·069
9	1421·5	2·426	2·411	2·420	2·426	3·096	3·141	3·094	3·101
10	1552·7	2·419	2·403	2·412	2·418	3·134	3·178	3·126	3·138
11	1680·3	2·413	2·395	2·404	2·411	3·171	3·216	3·160	3·176
12	1804·0	2·408	2·387	2·399	2·405	3·208	3·255	3·192	3·214
13	1924·0	2·402	2·381	2·390	2·399	3·242	3·295	3·225	3·250
14	2040·2	2·398	2·374	2·384	2·394	3·282	3·334	3·260	3·288
15	2152·3	2·395	2·368	2·380	2·389	3·325	3·373	3·288	3·323
16	2260·6	2·389	2·362	2·376	2·384	3·360	3·413	3·320	3·361
17	2364·9	2·385	2·358	2·372	2·380	3·401	3·452	3·350	3·399
18	2465·0	2·383	2·354	2·367	2·376	3·444	3·493	3·381	3·438
19	2561·1	2·379	2·349	2·363	2·373	3·486	3·538	3·414	3·479
20	2652·9	2·376	2·344	2·358	2·370	3·532	3·581	3·447	3·531
21	2740·6	2·374	2·340	2·355	2·368	3·575	3·627	3·477	3·579

h and c values throughout taken from Birge (1929 a), so that data are comparable with those published previously.

The values of ω_e, $\omega_e x_e$ and $\omega_e y_e$ appropriate to levels below and above $v'=37$ were evaluated by the method of least squares to give

$$21 \leqslant v' \leqslant 37 \qquad G(v')=172{\cdot}70(v'+\tfrac{1}{2})-2{\cdot}212(v'+\tfrac{1}{2})^2+4{\cdot}98\times10^{-3}(v'+\tfrac{1}{2})^3,$$

$$37 \leqslant v' \leqslant 48 \qquad G(v')=184{\cdot}64(v'+\tfrac{1}{2})-2{\cdot}857(v'+\tfrac{1}{2})^2+1{\cdot}373\times10^{-2}(v'+\tfrac{1}{2})^3.$$

If we assume that no further discontinuity occurs beyond $v'=37$, then we may use the values of ω_e, $\omega_e x_e$ and $\omega_e y_e$ for $v'>37$ to compute the dissociation energy by means of equation (13). The value obtained, 3808 cm.$^{-1}$, is in close agreement with the value 3814 cm.$^{-1}$ obtained by Brown using the graphical extrapolation method of Birge. The calculated value of $v'_{\max}$ is 51.

§ 9. ACKNOWLEDGMENTS

I am indebted to Messrs. R. H. Dalitz and A. K. Head for many helpfu suggestions.

REFERENCES

Birge, R. T., 1929 a. *Phys. Rev.*, Suppl. (*Rev. Mod. Phys.*), **1**, 1.
Birge, R. T., 1929 b. *Trans. Faraday Soc.*, **25**, 707.
Brown, W. G., 1931. *Phys. Rev.*, **38**, 1179.
Crawford, F. H. and Jorgensen, T., 1936. *Phys. Rev.*, **49**, 745.
Klein, O., 1932. *Z. Phys.*, **76**, 226.
Milne-Thomson, L. M., 1931. *Die elliptischen Funktionen von Jacobi* (Berlin).
Morse, P. M., 1929. *Phys. Rev.*, **34**, 57.
Oldenberg, O., 1929. *Z. Phys.*, **56**, 563.
Rosen, N. and Morse, P. M., 1932. *Phys. Rev.*, **42**, 210.
Rydberg, R., 1931. *Z. Phys.*, **73**, 376.
Rydberg, R., 1933. *Z. Phys.*, **80**, 514.
Whittaker, E. T. and Watson, G. N., 1940 a. *Modern Analysis* (Cambridge), p. 439 ; 1940 b. *Ibid.*, p. 451 ; 1940 c. *Ibid.*, p. 503.

NOTE ON THE INTERPRETATION OF THE VISIBLE ABSORPTION SPECTRUM OF BROMINE

By A. L. G. REES,
Division of Industrial Chemistry, Council of Scientific and
Industrial Research, Melbourne, Australia

MS. received 15 *January* 1947

ABSTRACT. The accurate construction of the lower parts of the potential-energy curves for the excited states involved in the transitions responsible for the visible absorption spectrum of bromine has enabled a choice to be made between two possible alternative interpretations.

The interpretation of the long-wavelength spectra of the halogen molecules has been discussed in detail by Mulliken (1940), who has found it possible to assign transitions to the various observed continua with certainty in all cases but bromine, where existing experimental and theoretical data do not allow a decision between two alternatives to be made. The possible interpretations of the two overlapping continua in the visible absorption of bromine vapour, designated A($\nu_{\max}=24\,300$ cm.$^{-1}$) and B($\nu_{\max}=20\,740$ cm.$^{-1}$) (Acton, Aickin and Bayliss, 1936), are:—

(i) A $^1\Sigma_g^+ \rightarrow {}^1\Pi_u$,
B $^1\Sigma_g^+ \rightarrow {}^3\Pi_{0^+u}$;
$^1\Sigma_g^+ \rightarrow {}^3\Pi_{1u}$ calculated to have $\nu_{\max}=18\,630$ cm.$^{-1}$ and to be relatively very weak.

(ii) A $^1\Sigma_g^+ \rightarrow {}^1\Pi_u$,
B $^1\Sigma_g^+ \rightarrow {}^3\Pi_{1u}$;
$^1\Sigma_g^+ \rightarrow {}^3\Pi_{0^+u}$ calculated to occur at $l\nu_{\max}=22\,800$ cm.$^{-1}$ and to be approximately equal to intensity in $^1\Sigma_g^+ \rightarrow {}^3\Pi_{1u}$.

Franck-Condon evidence (Bayliss, 1937; Darbyshire, 1937) appears to be in favour of interpretation (i). The evidence is not conclusive, since the lower part of the $^3\Pi_{0^+u}$ curve, fitted by Bayliss to the upper segment of the curve obtained from the B continuum, was constructed by the use of the Morse function, which deviates from the true curve at high vibrational levels. Moreover, the weak continuum arising from the $^1\Sigma_g^+ \rightarrow {}^3\Pi_{1u}$ transition is expected in the region of the band absorption and has not been detected experimentally. On the other hand, certain of the effects observed in bromine solutions (Aickin, Bayliss and Rees, 1938) could be interpreted in terms of (ii) rather than (i) (Bayliss and Rees, 1939). As Mulliken (1940) has pointed out, an unambiguous decision could be made by an accurate construction of the lower parts of the $^3\Pi_{0^+u}$ and $^3\Pi_{1u}$ curves from the existing band-spectroscopic data. Accurate potential-energy curves may now be constructed by analytical methods described in the preceding paper (Rees, 1947), and the laborious graphical procedure of the Rydberg-Klein method avoided.

As there were some points of disagreement in the independent evaluation of the spectroscopic constants from the band data by Brown (1931) and Darbyshire (1937) respectively, the derivation of the constants has been re-examined. For $^3\Pi_{0^+u}$, a least-squares calculation shows that a cubic in $(v'+\frac{1}{2})$, with numerical values of the coefficients as given by Brown, is necessary to describe the data up to $v'=21$, although Darbyshire claims that a quadratic is adequate up to $v'=15$. Darbyshire's value of E_e for this state appears to be in error by the quantity $2G'(0)$; the correct value is 15749 cm.$^{-1}$ For $^3\Pi_{1u}$ the data lead to a value for E_e of 13856 cm.$^{-1}$ and not 13814 cm.$^{-1}$ as quoted by Darbyshire. These values are important in fixing the $^3\Pi_{1u} - {}^3\Pi_{0^+u}$ interval (now 1893 cm.$^{-1}$ and *not* 2104 cm.$^{-1}$) and the relationship of these states to the ground state potential-energy curve.

Using these data for ^{79}Br^{81}Br, the curve for $^3\Pi_{0^+u}$ was constrcuted up to $v'=22$ (within 170 cm.$^{-1}$ of D_e' for this state) and for $^3\Pi_{1u}$ up to $v'=22$ (within 150 cm.$^{-1}$ of D' for this state) by method (iv) of the preceding paper (Rees, 1947). For $^3\Pi_{0^+u}$ the curve up to $v'=21$ was calculated also by method (i), employing the analytical expressions equivalent to the Rydberg-Klein graphical method. These curves are plotted in the figure, together with the upper segments derived by Bayliss (1937) from the analysis of the continuous absorption data. The lower part of the $^3\Pi_{0^+u}$ curve extrapolates perfectly (within 0·005 A.) on to Bayliss's upper segment corresponding to the B continuum, whereas the corresponding Morse curve is 0·03 A. in error. A reasonable extrapolation of the $^3\Pi_{1u}$ curve would lie about 1900 cm^{-1} below this curve at r_e'' (=2·284 A.) and could not be extrapolated on to the segment labelled B. The curve for the repulsive $^1\Pi_u$ state has been drawn to cross $^3\Pi_{0^+u}$ in the region of the 3rd and 4th vibrational levels, as suggested previously by Bayliss and Rees (1939).

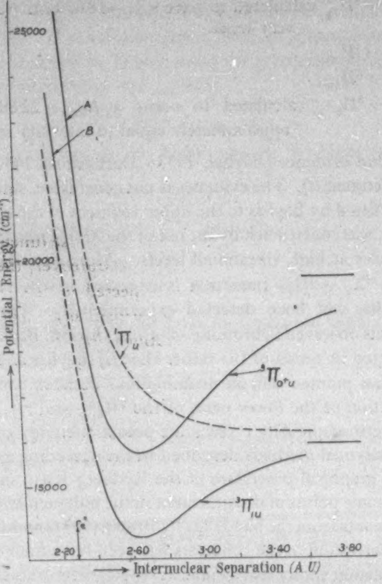


Figure 1. Potential energy curves for the upper states involved in the visible absorption spectrum of Br_2.

This evidence is consistent only with interpretation (i). Moreover, it directly lends strong support to the Franck-Condon principle, as it seems very improbable that the continuity of $U'(r)$ and its first derivative at D_e' should be fortuitous.

REFERENCES

Acton, A. P., Aickin, R. G. and Bayliss, N. S., 1936. *J. Chem. Phys.*, **4**, 474.
Aickin, R. G., Bayliss, N. S. and Rees, A. L. G., 1938. *Proc. Roy. Soc.*, A, **169**, 234.
Bayliss, N. S., 1937. *Proc. Roy. Soc.* A, **158**, 551.
Bayliss, N. S. and Rees, A. L. G., 1939. *J. Chem. Phys.*, **7**, 854.
Brown, W. G., 1931. *Phys. Rev.*, **38**, 1179.
Darbyshire, O., 1937. *Proc. Roy. Soc.*, A, **159**, 93.
Mulliken, R. S., 1940. *Phys. Rev.*, **57**, 500.
Rees, A. L. G., 1947. *Proc. Phys. Soc.*, **59**, 998.

THE EFFECTIVE PERMITTIVITY OF TWO-PHASE SYSTEMS

By D. F. RUSHMAN and M. A. STRIVENS,
Material Research Laboratories, Philips Electrical Ltd., Mitcham

MS. received 7 *February* 1947

ABSTRACT. The effective permittivity of samples of barium titanate with porosities up to 40% v/v. has been determined, and an explanation given in terms of Wiener's mixture law. It is shown that the values of the form factor in this law can only be predicted with certainty over the ranges 0–30% and 70–100% porosity, and takes up intermediate values in the 30–70% porosity region. The observed values of ϵ are in accordance with this hypothesis.

---

§ 1. INTRODUCTION

It is often required to find the true permittivity of a substance when it is a component of a heterogeneous mixture whose effective permittivity can be measured and where the permittivities of the other components are known. It is also useful for practical applications to be able to calculate the effective permittivity from a knowledge of the volume fractions and the permittivities of the components.

This problem, which also occurs in the magnetic case of the effective permeability of iron-dust cores, etc., has recently been brought into prominence owing to the study of the ferroelectric behaviour of barium titanate, which has only been prepared so far in the form of sintered polycrystalline masses, which are invariably porous to a certain extent.

§ 2. EXPERIMENTAL

Barium metatitanate, $BaTiO_3$, was prepared by milling together equimolecular proportions of the purest available barium carbonate and titanium dioxide, pressing the mixture into blocks and prefiring it at 1250° c. The product was then crushed and re-milled. A chemical analysis at this stage showed the total amount of impurities to be less than 0·2%, with 49·12 mol % BaO and 50·86 mol % TiO_2.

Using this material samples were made by pressing the powder in a die to give discs about 1 cm. × 3 mm. and firing at temperatures ranging between 1000° c. and 1350° c.

This resulted in a series of samples having different porosities and hence different permittivities. The porosity was determined in each case from the expression

$$P = \frac{\rho_x - \rho_A}{\rho_x},$$

where $\rho_x = 6{\cdot}08$ g/cc. is the x-ray density and ρ_A is the apparent (or bulk) density. The latter was determined directly by grinding the samples to a regular shape, weighing them, and measuring them with a micrometer.

Electrodes were applied by evaporation of silver *in vacuo* on to the faces of the discs, and the capacities were measured *in vacuo* at 1·6 Mc/sec. and 20° c. by a substitution method.

The results given in table 1 are shown graphically in figure 1.

Table 1

Sample No.	Firing temp. (° c.)	Density (g/cc.)	Porosity vol. (%)	ϵ_m
1	1130	3·56	41·2	458
2		3·59	40·7	443
3		3·52	41·8	439
4	1150	3·61	40·3	491
5		3·60	40·5	520
6		3·61	40·3	473
7		3·67	39·3	508
8		3·64	39·8	486
9	1200	4·08	32·6	894
10		4·06	32·9	918
11		4·04	33·2	897
12	1250	4·29	29·0	956
13		4·26	29·6	930
14		4·18	31·0	956
15		4·25	29·8	922
16		4·28	29·2	957
17		4·12	31·9	956
18	1300	4·92	18·7	1176
19		4·70	22·3	1160
20		4·68	22·6	1084
21	1350	5·37	11·2	1366
22		5·41	10·5	1381
23		5·45	9·9	1408
24		5·43	10·3	1423
25		5·55	8·3	1447
26		5·45	9·9	1418

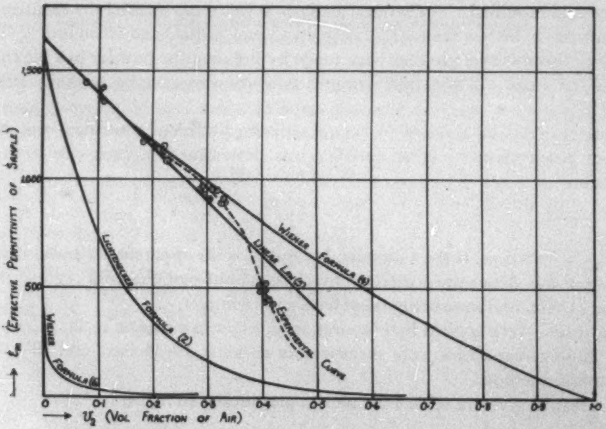


Figure 1.

§ 3. DISCUSSION

Numerous mixture laws are recorded in the literature, some of which have an empirical or semi-empirical basis, others, however, having a more theoretical significance. It is evident from the large quantity of recorded experimental data, and on general grounds also, that several factors, apart from the relative volumes and the permittivities of the constituents, must be accounted for in a general mixture law. The chief of these is the shape or form of the different phases as distinct from their internal structure, which is assumed to be continuous and homogenous and, in particular, which are the continuous and which are the discontinuous phases.

For polarizable spheres in a vacuum the Clausius-Mosotti formula has a sound theoretical basis and leads to the mixture law

$$\frac{\epsilon_m - 1}{\epsilon_m + 2} = \sum_i V_i \frac{\epsilon_i - 1}{\epsilon_i + 2},$$

where V_i is the volume fraction of the constituent of permittivity ϵ_i, and ϵ_m is the effective permittivity of the mixture.

This formula has the support of considerable experimental data, in particular relating to non-polar gases and certain liquids, and is quite valid provided that the concentration of spheres is not too high and that they have neither a random distribution or a cubic symmetry.

The Clausius-Mosotti formula may be generalized to include the case of polarizable spheres embedded in a medium of permittivity ϵ_0. The factor $(\epsilon - 1)/(\epsilon + 2)$ then becomes $(\epsilon - 1)/(\epsilon + 2\epsilon_0)$, and the formula will still be valid within the limits imposed by the model.

Wiener (1904 and 1910) has generalized the formula further and states it in the following form:

$$\frac{\epsilon_m - 1}{\epsilon_m + u} = \sum_i V_i \frac{\epsilon_i - 1}{\epsilon_i + u}, \qquad \ldots\ldots(1)$$

where u is a form factor governed by the shape of the phase present. Table 2 shows the values of u for certain 2-phase systems together with the form which equation (1) assumes. It will be noted that $u = 2\epsilon_0$ for spheres embedded in a medium of permittivity ϵ_0.

A commonly used law is that of Lichtenecker (1926), viz.,

$$\log \epsilon_m = \sum_i V_i \log \epsilon_i, \qquad \ldots\ldots(2)$$

which is derived as an intermediate form of the series and parallel combination laws for dielectrics. Although its theoretical significance is slight, it has met with some success, but only in cases where the difference between the permittivities of the constituents have been small.

In our case $\epsilon_1 \gg \epsilon_2$, where the subscripts 1 and 2 refer to $BaTiO_3$ and air respectively, and reference to figure 1 shows that the observed values show no agreement whatsoever with the Lichtenecker formula.

The observed permittivity decreases almost linearly with increasing porosity until a value of about 35% porosity is reached, after which it falls rapidly. This rather unexpected behaviour is best understood in terms of the Wiener formula (1).

If we regard the pores in barium titanate as spherical (to a first approximation), we may put $u = 2\epsilon_1$, where ϵ_1 is the true permittivity of $BaTiO_3$.

Table 2

Arrangement	u-factor	Formula
Layers perpendicular to the lines of force	0	$\frac{1}{\epsilon_m}=\frac{V_1}{\epsilon_1}+\frac{V_2}{\epsilon_2}$
Cylinders of I in II to lines of force	ϵ_2	$\frac{\epsilon_m-\epsilon_2}{\epsilon_m+\epsilon_2}=V_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+\epsilon_2}$
Spheres of I in II	$2\epsilon_2$	$\frac{\epsilon_m-\epsilon_2}{\epsilon_m+2\epsilon_2}=V_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+2\epsilon_2}$
Cylinders of II in I to lines of force	ϵ_1	$\frac{\epsilon_m-\epsilon_1}{\epsilon_m+\epsilon_1}=V_2\frac{\epsilon_2-\epsilon_1}{\epsilon_2+\epsilon_1}$
Spheres of II in I	$2\epsilon_1$	$\frac{\epsilon_m-\epsilon_1}{\epsilon_m+2\epsilon_1}=V_2\frac{\epsilon_2-\epsilon_1}{\epsilon_2+2\epsilon_1}$
Layers or discs parallel to lines of force	∞	$\epsilon_m=V_1\epsilon_1+V_2\epsilon_2$

Formula (1), written for a two-phase system, now takes the form

$$\frac{\epsilon_1-\epsilon_m}{\epsilon_m+2\epsilon_1}=V_2\frac{\epsilon_1-\epsilon_2}{\epsilon_2+2\epsilon_1}, \qquad \text{......(3)}$$

where $V_2=(1-V_1)$ is the volume fraction of pores.

In this form Wiener's law should be applicable to any system of spheres dispersed in a continuous medium, but in our case it can be simplified since $\epsilon_1 \gg \epsilon_2$. We have, therefore,

$$\frac{\epsilon_1-\epsilon_m}{\epsilon_m+2\epsilon_1} \simeq \frac{V_2}{2},$$

or, that is,

$$\epsilon_m \simeq \frac{2\epsilon_1(1-V_2)}{(2+V_2)}=\epsilon_1\left[1+3\sum_{n=1}^{\infty}\left(-\frac{V_2}{2}\right)^n\right]. \qquad \text{......(4)}$$

In cases where the total porosity is small, the power series may be taken to only the first term with considerable accuracy, i. e.

$$\epsilon_m \simeq \epsilon_1(1-\tfrac{3}{2}V_2). \qquad \text{......(5)}$$

This approximate form, which has been used elsewhere by us to correct for porosity in high permittivity materials (Rushman and Strivens 1946), has also been derived as a limiting case by Polder and van Santen (1946). Figure 1 shows that the experimental points give good agreement with equation (4) for values of V_2 less than 0·35, and that the approximation (5) is quite valid for extrapolation to zero porosity in this region. We have, however, used equation (4) for our extrapolation and found the true permittivity of this sample of $BaTiO_3$ to be 1650.

The rapid drop in the observed value of ϵ_m in the region of 35% porosity is due to several interdependent factors.

Firstly, there is a limit to the packing density of the air pores. Assuming that they are equal in size and arranged in cubic close-packed symmetry, the limiting porosity beyond which they coalesce is equal to $\pi/3\sqrt{2}$ or about 70%. Below this limiting porosity the Wiener formula (4) with $u = 2\epsilon_1$ would apply if the concentration of spheres were not so high that the field in the not too close proximity of each pore ceases to be equivalent to the average or smoothed-out field, which is one of the basic assumptions behind the Clausiu-Mosotti formula. One would expect for this reason alone a breakdown in the Wiener formula (4).

But secondly, it is extremely probable that the air pores have a random distribution and are unequal in size, so that they will begin to coalesce long before the limiting porosity for cubic close-packing, yielding pores shaped like simple or branched filaments, for which the form factor, u, cannot be calculated for the general case.

A similar picture holds if we reverse the phases and consider particles of barium titanate dispersed in air. The relevant mixture law is obtained by interchanging subscripts in equation (3), giving

$$\frac{\epsilon_2 - \epsilon_m}{\epsilon_m + 2\epsilon_2} = V_1 \frac{\epsilon_2 - \epsilon_1}{\epsilon_1 + 2\epsilon_2},$$

which for $\epsilon_2 = 1$ and $\epsilon_1 \gg \epsilon_2$ reduces to

$$\epsilon_m \simeq \epsilon_1 \frac{(3 - 2V_2)}{3 + V_2\epsilon_1}. \qquad \text{......(6)}$$

As before, a limiting packing fraction ($V_1 \simeq 0{\cdot}70$) exists beyond which the barium titanate cannot be completely discontinuous. In the intermediate region extending approximately from 30% to 70% porosity we expect, therefore, to find either or both of the phases continuous or partially continuous.

The model assumed for sintered polycrystals of barium titanate is, therefore, one of spherical air pores dispersed in a continuous medium of barium titanate for porosities up to about 35%, for which region $u = 2\epsilon_1$, and for porosities from 35% to about 45% it is assumed that the barium titanate is wholly continuous and the air partially continuous, the u factor being indeterminate. Beyond about 45% porosity the barium titanate is discontinuous and the air continuous since the particles cannot be sintered together at such high porosity by this method. The u-factor in this region, assuming spherical particles not too close together, is $2\epsilon_2$ as stated in table 2. The "high porosity" formula (6) should, however, be applicable to systems such as $BaTiO_3$/wax, in which the low-permittivity wax forms the continuous phase.

It is not possible to compute a generalized u-factor for the region 30% to 70% porosity and hence to deduce a mixture law, but if normal methods of preparation are used one would expect a smooth transition between the Wiener formulae (4) and (6), which in turn implies smooth transition between the corresponding u-factors, $2\epsilon_1$ and $2\epsilon_2$.

Little can be said about the precise nature of the variation in u over the intermediate region, as it will depend on the exact conditions of preparation of the material, e.g. the density distribution within the compacted material before sintering, and more especially on the physical forces acting on the particles during the sintering itself process. In this region there would appear to be little hope of deducing a rigorous mixture law, and the only formula having any practical utility would be one giving an empirical relationship between the *u*-factor and composition.

ACKNOWLEDGMENTS

The authors wish to thank Mr. J. A. M. van Moll and the directors of Philips Electrical Ltd. (formerly Philips Lamps Ltd.) for permission to publish this paper.

REFERENCES

Lichtenecker, 1926. *Phys. Z.*, **27**, 115.
Polder and van Santen, 1946. *Physica*, **12**, 257.
Rushman and Strivens, 1946. *Trans. Faraday Soc.*, **42** A, 231.
Wiener, 1904. *Phys. Z.*, **5**, 332.
Wiener, 1909. *Leipzig Ber.*, **61**, 113; 1910. *Ibid.*, **62**, 256.

DIFFUSION PUMPS: A CRITICAL DISCUSSION OF EXISTING THEORIES

By D. G. AVERY* and R. WITTY,
Metropolitan-Vickers Electrical Co. Ltd., Manchester, 17

MS. received 8 *May* 1947; *read at Science Meeting on* 31 *October* 1947

ABSTRACT. The original work of Gaede and others on diffusion pumps is briefly described. A critical discussion of this work leads to the formulation of a more complete theory of the action of the diffusion pump, and this is used to explain the practical characteristics of a simple form of modern diffusion pump.

§ 1. INTRODUCTION

THE main contributions to the theory of the diffusion pump between 1915 and 1923 were those made by Gaede (1915, 1923), Langmuir (1916) and Crawford (1917). Since that time several critical surveys of these theories (see General Bibliography) have been made, but it is felt that none of them have reached definite conclusions. The purpose of this paper is to correlate the previous work mentioned above and then put forward a theory of the diffusion pump which will satisfactorily explain the known phenomena of the pumps. In this paper considerations are limited to cases where the working pressure of the pump is less than 10^{-3} mm. of mercury.

§ 2. ORIGINAL THEORIES

(a) *Gaede's theory of the diffusion pump*

The original work which led to the development of the Gaede air-diffusion pump (Gaede, 1915) was concerned with diffusion processes due to gas and vapour counter-flowing through a tube under the conditions shown in figure 1.

---

* Now at Royal Holloway College.

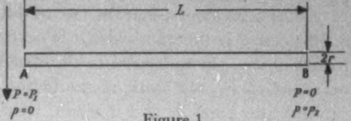


Figure 1.

Suppose that at the two ends A and B of the tube, of radius r and length L, the vapour pressure P has the values P_1 and zero, and the gas pressure p has values zero and p_2, respectively. When these conditions are satisfied (for example by creating a gas-free vapour stream in the direction of the arrow at A, and placing a condenser at B), then the volume of gas V flowing per second from B to A is given by

$$V=\frac{1}{L}\cdot\frac{\pi r^3}{2v_1}\cdot\exp\frac{-rP_1}{1520Dv_2}, \qquad \ldots\ldots(1)$$

where v_1 and v_2 are the coefficients of external friction of gas and vapour respectively, and D is the coefficient of diffusion for the gas and vapour concerned. It should be noted that V is greatest when rP_1 is least and L is small. Thus if the vapour pressure P_1 is high, the radius r must be very small, and the volume V may be increased by using a number of similar tubes in parallel; alternatively, if the vapour pressure P_1 is very low, then r may be large. Since at low gas pressures for gas and vapour counter-flowing in a tube, the vapour pressure P_1 completely determines the value of the mean free path, λ, of the gas molecules in the vapour, it follows that the volume V depends on λ.

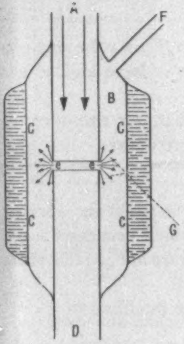


Figure 2.

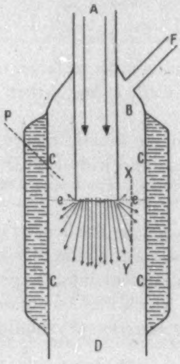


Figure 3.

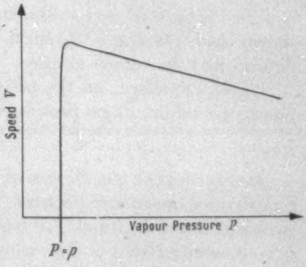


Figure 4.

Gaede (1915, 1923) used these ideas as a basis for the construction of the pumps of the type shown in figure 2. In these pumps the diffusion processes take place through the gap e. A stream of gas-free vapour passes in the direction shown from A to D past this gap and is condensed at D in a space already evacuated by a backing pump. Some of this vapour passes out through e into the space B, emerging in a brush-like formation to be condensed at the cooled walls C. Air from the vessel to be exhausted enters the system at F and, after diffusing through this cloud of vapour, enters the stream at e. Thus the conditions prescribed in the previous paragraph are satisfied.

It has been shown that the rate of diffusion depends on the product rP_1, where in this case r is the gap-width. Therefore, since in the region G the mean free path λ of the gas molecules in the vapour issuing from the gap is inversely proportional to P_1, the value of the factor r/λ controls the rate at which gas passes through the pump. If λ is small, then the probability that a gas molecule will penetrate through the vapour from G and enter the stream at e is small, and, in the limiting case, the issuing vapour sweeps the gas back from the region of the gap. On the other hand, if λ is large and r/λ small, then the gas easily enters the stream, suffering very few adverse collisions with vapour molecules before reaching e. If P_1 is made smaller, and λ in consequence larger, then it follows that the gap may be made larger and an increase in speed obtained. An upper limit is placed on the free path length since, if this is very great, and the vapour pressure correspondingly very low, the gas that has passed through the gap is no longer carried to the fore-vacuum at D. Gaede stated as a condition for optimum working that the mean free path λ should be of the same order of magnitude as the gap-width r.

The pressure exerted by the vapour at G on the air in the vessel B, termed the *diffusion back pressure*, is given by

$$\Delta p = p_2\left(1 - \exp \frac{-rP_1}{1520Dv_2}\right). \qquad \ldots\ldots(2)$$

The exponential term in this equation must be large, i.e. rP_1 must be small, as shown above, for Δp to be small. If rP_1 is large, then Δp is large and no gas diffuses into the vapour stream.

Gaede calculated, on the basis of the kinetic theory, that, for this form of pump, the volume of gas passing per second from F to D is given by

$$V = \alpha q / 2\pi\rho_1. \qquad \ldots\ldots(3)$$

In this equation q is the area of the gap, ρ_1 the density of the gas at the working temperature under unit pressure, and α a function of r/λ which decreases as r/λ increases, or as λ decreases. V is independent of p_2, the pressure at F, so that there is no theoretical limit to the vacuum which may be obtained.

Gaede further showed (1923) that equation (3) should be modified to

$$V = k\frac{\alpha q}{2\pi\rho_1}, \qquad \ldots\ldots(4)$$

where k is a constant taking into account the fact that the vapour stream may not be able to remove all the gas arriving at e. He considered that the layers of vapour nearest the gap e become saturated with gas and can no longer entrain all the gas arriving. He claimed that the amount of gas which any stream can "absorb" is greatest when the individual points of the vapour stream remain exposed to the gas for the shortest possible time, so that, for a given vapour stream, k is large when r is small, or, for a given r, k can be increased by increasing the velocity of the stream, e.g. by using driving nozzles. In the case of pumps of the type shown in figure 3, with the diffusion gap at e, where the density of the stream increases rapidly with the depth of penetration along a line such as XY, Gaede suggested that the surface layers may become more easily saturated and the value of k correspondingly less.

Equation (4) gives the volume of gas passing per second from F to D (figure 2); it is not, however, a true equation for the pumping speed. Gaede quoted the complete equation as

$$V=\frac{q}{2\pi p_1}\left(k\alpha-\beta\frac{p_1}{p_2}\right), \qquad \ldots\ldots(5)$$

where p_1 is the pressure existing in D, the second term accounting for gas back-diffusing from D to F. The quantity β decreases rapidly with increasing vapour pressure and stream speed, and so, for any particular value of p_1, the speed is independent of this value once P_1 has exceeded a value such that $\beta=0$. On the other hand, if P_1 is equal to or less than p_1, then the stream can no longer hold back the pressure in D and the pump ceases to function.

If the speed is plotted aainst the vapour pressure, the curve shown in figure 4 is obtained. The pump has zero, or negative, speed so long as $P_1<p_1$, and immediately rises to a maximum as soon as P_1 equals p_1. Thereafter the speed slowly declines with increasing P_1, since λ and α both decrease.

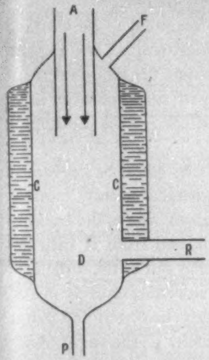


Figure 5.

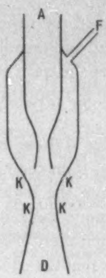


Figure 6.

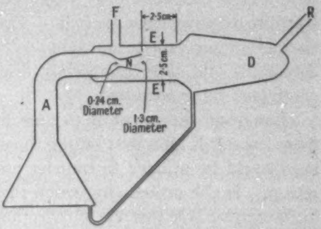


Figure 7.

Gaede concluded the theoretical discussion by pointing out that if the speed of the stream is increased, the speed of pumping is increased, not only due to the increase in *k*, already mentioned, but also, especially in the type of pump shown in figure 3, due to the increase in α as a result of the reduction of the chances of a vapour molecule "back-streaming" through the gap e into the space B.

Several interesting experimental confirmations of Gaede's theory have been made by Molthan (1925, 1926).

(*b*) *Langmuir's theory of the condensation pump*

A typical Langmuir pump (Langmuir, 1916) is shown in figure 5. Mercury vapour passes down the tube A into the larger tube D, the walls of which are cooled by the condenser C. The vapour, on leaving the tube A, spreads out and is condensed on the walls, whence it returns to the boiler by way of the tube P. The high-vacuum side of the pump is at F, the backing pump being connected at R.

Langmuir considered the action of all high-vacuum vapour pumps to be divided into two separate processes, viz.,

(1) the process by which the gas enters the blast of vapour; and

(2) the process by which the blast carries the gas along into a condensing chamber from which it cannot return to the vessel being exhausted.

The second of these processes is practically the same for Gaede's diffusion pumps, steam-ejector pumps, and the pump under consideration. Langmuir considered that the effectiveness of the Gaede pump lay in the success of the second process, and that the limitations were those imposed by the first, that is by the restrictions on the size of the diffusion orifice.

Langmuir proceeded to consider ejector pumps; in such pumps the gas is drawn into the vapour because the pressure in the jet at the point where the gas enters it is lower than the pressure of the surrounding atmosphere, that is, these pumps utilize the Bernouilli effect. This effect cannot be directly utilized in the production of high vacua, since the pressure in the jet must always be considerable and the jet will expand laterally, as predicted by the kinetic theory.

If an ejector pump such as that shown in figure 6 is assumed to be working with a high vacuum on the intake side, then the vapour blast from A will expand laterally to strike the walls at K. The wall will rapidly assume the temperature of the vapour and vapour molecules will leave the walls at K with high velocities in all directions and thus completely destroy the pumping effect. This Langmuir confirmed by experiment.

Langmuir then went on to consider that if these molecules could be prevented from leaving K after first hitting the walls, then their downward-velocity component could be utilized to transfer gas from the high-vacuum side to the backing pump. If the process by which the vapour molecules passed from K into the high-vacuum side was one of reflection, then there was no way out of the difficulty, but Langmuir's previous work had shown that the effect was almost entirely one of condensation followed by random re-emission. Accordingly, he concluded that if the walls of the pump were kept cool, as in figure 5, by the cooler C, then a pumping action could be obtained.

This was borne out by experiment, and Langmuir (1916) summarized the action of his pump as follows:—"In these pumps a blast of mercury vapour carries the gas into a condenser. This action is similar to that in a steam ejector and in a Gaede diffusion pump. The method by which the gas is brought into the mercury-vapour blast in the condensation pump is based on a new principle which is essentially different from that employed in the steam ejector or Gaede diffusion pump. In the new pumps the gas to be exhausted is caught by the blast of vapour and is forced by gas friction to travel along a cooled surface. By maintaining this surface at such a low temperature that the condensed mercury does not re-evaporate at an appreciable rate, it is possible to keep the mercury vapour from escaping into the vessel being exhausted. The action of this pump therefore depends primarily upon the fact that all the atoms of mercury striking a mercury-covered surface are condensed (no matter what the temperature), instead of even a fraction of them being reflected from the surface. It is for this reason that the term 'condensation pump' is proposed."

From these considerations it can be seen that such a pump should have no limiting vacuum, since a limit could only be imposed by the back-diffusion of gas through the stream, and calculations show that the chance of a gas molecule finding its way back through the stream in a typical pump is only 1 in 10^{20}.

(c) *Crawford's theory of the parallel jet pump*

Crawford (1917) considered that the primary obstacle to the successful operation of vapour pumps was the fact that the vapour stream dispersed when surrounded by a high vacuum, and did not satisfactorily entrain the gas from the vessel to be exhausted.

He considered that if a jet could be produced, the molecules of which had equal and parallel velocities, one with another, then such a jet of vapour should be collision-free and should not materially disperse in a high vacuum. Also it would be a very efficient pumping agent since gas could readily enter the stream in a direction perpendicular to its axis, but gas molecules once in the stream would be carried along with it in the direction of its motion. He further considered that, even if collisions did occur in such a jet, the stream would not necessarily be dispersed since only the direction of relative velocities, and not the velocity of the centre of gravity of the whole mass, would be altered by such collisions.

These considerations led Crawford to design pumps of the type shown in figure 7. The form of nozzle N was settled by the considerations of steam-engineering practice, the point of minimum pressure and maximum velocity existing in the diverging portion of the nozzle. With such a pump, Crawford stated that no diffusion slit or condensing surface was necessary. The tube E could be artificially heated without seriously affecting the performance, this suggesting that the jet itself re-entrained and expelled vapour molecules diffusely emitted from the walls. He calculated that in a typical jet a vapour molecule would suffer, on an average, six collisions between leaving the nozzle and being condensed in D.

If the jet-density exceeded a well-defined limit, he found that the pump ceased to work, and he explained this by saying that "this limit is established by the density of the dispersing fringe, which is probably proportional to the density of the jet and occurs at the point where the mean free path of the gas molecules entering the fringe becomes less than the total depth of the fringe".

§ 3. DESIGN DETAILS

(a) *Gaede*

Gaede's first designs of pumps (1915) were all based directly on the fundamental arrangement shown in figure 2. A thermometer was usually inserted to measure the temperature of the vapour, and hence the vapour pressure, at some point near the gap, since the heating adjustment, for the correct value of P, was very critical (cf. figure 4). With these pumps he obtained an optimum speed of 80 cc./sec. with air.

Following on from the theoretical considerations discussed, he added a driving nozzle to increase the velocity of the vapour stream (1923), giving the arrangement shown in figure 8. With this pump he obtained a speed of 0·25 litres/sec., using the same gap-width as in the first pump mentioned above. Since the effective mean free path of the vapour molecules in the stream had been increased by the use of the nozzle, it was possible to increase the gap-width and, using an annular nozzle ring shaped in section, to give a high-velocity stream, he obtained a speed of 1·5 litres/sec. using the same tube diameters as those in figure 8. A larger pump, in which the tube B had an internal diameter of 5 cm., gave speeds of up to 6 litres/sec. He found that when a nozzle was inserted, the adjustment of vapour temperature had no longer such a critical effect on the speed.

Gaede's latest type (1923) was a multi-stage pump developed from pumps of the type shown in figure 3, again utilizing a driving nozzle. The first stage of the pump had a nozzle form as shown in figure 9 (*a*), the diffusion orifice again being at e. This stage was backed by two or more stages of the general type shown in figure 9 (*b*). With such a pump Gaede obtained speeds of 60 litres/sec., with a value of $k\alpha$ in equations (4) and (5) of 0·4.

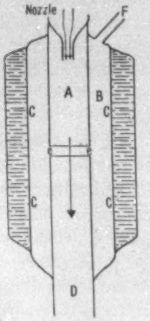
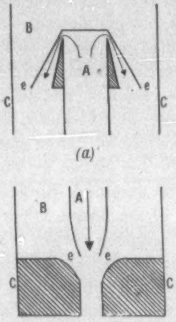


Figure 8.



(a)

(b)

Figure 9.

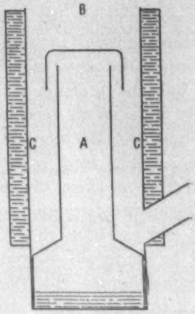


Figure 10.

In all cases, when working with air, he found that the speed and ultimate pressure were independent of the backing pressure below a certain critical value of that pressure.

(b) *Langmuir*

The first design of condensation pump has already been shown in figure 5; it should be noted that this system is identical with the system shown in figure 3. In these pumps, as in the later Gaede pumps, the adjustment of temperature was by no means critical. Langmuir observed critical-backing pressure effects, and further found that the value of the critical-backing pressure increased rapidly with increasing heat input.

Langmuir's major contribution from the design point of view was the construction of all-metal vertical pumps of the type shown in figure 10. Using one of these pumps, with a casing internal diameter of 7 cm., he obtained a speed of 3 litres/sec.

(c) *Crawford*

The design of Crawford's pump (1917) was discussed in §2(c), the nozzle being designed from steam-engineering principles. A dimensional drawing of one of his pumps is shown in figure 7. The curves of speed against boiler pressure show a definite, though not critical, optimum boiler pressure, giving, for a typical pump, a speed of 640 cc./sec. Other pumps had speeds as high as 1·2 litres/sec.*

§ 4. CRITICAL DISCUSSION

From the theories developed in the preceding two sections it appears possible to give a fairly complete description of the action of these types of high-vacuum vapour pumps. It can be seen that a successful pump must be constructed so as to fulfil satisfactorily two prime conditions; first, the gas on the fine side must have easy access to the pumping stream, so that it may be removed quickly to the backing side; secondly, the pump must act as an efficient pressure seal so as to give as low a pressure as possible on the fine side whilst operating against as high a backing pressure as possible. Following Langmuir, then, the action of a vapour pump may be divided into two parts:

(1) the entry of the gas into the stream;
(2) the removal of the gas to the backing side.

Back diffusion, critical backing pressure

The second of these processes is practically the same in all pumps. Once the incoming gas has been entrained in the stream, it is forced to move in the direction of the stream towards the backing side by large numbers of collisions with the comparatively dense, fast-moving, heavy vapour-molecules. These same collision processes prevent almost all the gas in the backing side from forcing its way back into regions of lower pressure. In all pumps the operating characteristics show that for heavy gases, and at all values of the backing pressure below a certain critical value, only a very small proportion of the gas on the backing side "back-diffuses" into the fine side space. Above this critical backing pressure the gas pressure on the backing side is greater than the pressure exerted by the stream, and the stream is forced back, with a consequent destruction of the pressure-sealing effect.

In the case of light gases greater amounts of gas back-diffuse at backing pressures below the critical, and since for any gas the amount back-diffusing (at pressures below the critical) is clearly dependent on that pressure, for light gases the critical effect is less marked.

Fundamental condition for efficient pumping

Process (1) is concerned with the entry of the gas into the pumping stream, that is with the passage of the gas from a point such as P, figure 3, into the pumping stream through the gap e. Now in any pump a certain number of vapour molecules will escape from the stream and proceed in the direction from e to P, and any gas molecule entering the stream has to pass through a "cloud" of these "backstreaming" vapour molecules, which are moving in straight-line paths from points in the pumping stream. Obviously the gas will enter the stream most readily ***when the mean free path of the gas molecules in the cloud of backstreaming vapour is as large as***

---

* In comparing the performances of these various types of pump it would be more instructive to give the values of $k\alpha$, since the areas q of the several pumps are not the same. Unfortunately, however, there are not sufficient published data to enable this to be done.

possible. Under these conditions, the number of collisions suffered by a gas molecule with the vapour is a minimum, which is desirable since each collision will result in the gas molecule largely assuming the direction of the velocity of the backstreaming molecule. *Therefore, for process* (1) *to be carried out efficiently in any pump, the backstreaming must be reduced to a minimum.*

Causes and elimination of backstreaming

Backstreaming is caused by collisions between the vapour molecules in the pumping stream occurring whilst the stream is exposed to the fine-side space, resulting in vapour molecules leaving the stream in directions contrary to the pumping direction. The density of the backstreaming molecules is proportional to the number of collisions so occurring, and, therefore, for a given speed of stream and size of diffusion orifice, it is proportional to the pressure and, therefore, the temperature of the molecules in the stream. *The only important differences between the various types of pump considered are those between the methods employed to reduce the backstreaming of the vapour molecules to a minimum.*

In Gaede's pump, figure 2, backstreaming was reduced by adjusting the temperature (and thus the density) of the stream and the width of the slit, so that the number of collisions occurring between the molecules of the stream whilst it was exposed to the fine side was a minimum. A lower limit was placed on the stream density by the considerations of process (2). When the speed of the vapour stream was artificially increased, the time during which any portion of the stream was exposed was reduced, and hence it was possible to increase the slit-width without fear of increased backstreaming.

In Langmuir's pump of the type shown in figure 5, backstreaming was reduced by constructing the system so that a vapour molecule had to turn through 180° to the direction of motion of the stream before it was proceeding directly contrary to the pumping direction, as opposed to 90° in the Gaede pump, and, since the walls of these pumps were vigorously cooled, there was very little random re-emission of vapour molecules from the walls after they had once passed across the gap. Some control on the number of collisions occurring in the stream whilst it was exposed to the fine side, had, of course, still to be maintained by the correct adjustment of the vapour temperature, but since any collision was less likely to have a serious effect on the action of the pump, this adjustment was not nearly so critical as in the Gaede pump.

Crawford constructed his pump so that the stream was directed in the pumping direction and so that the vapour molecules should have high velocities with near parallel directions. That the use of a high-velocity stream reduces the chance of a molecule backstreaming follows from the fact that for a molecule to stream back it must have a thermal velocity greater than the streaming velocity. Thus, if the streaming velocity is greater, fewer molecules will have thermal velocities greater than the streaming velocity. Gaede realized this, and, when he increased the speed of the stream by the use of driving nozzles, he was able to use a wider slit and obtain greater speeds, since the vapour molecules streamed back less readily.

Drift vs condensation

Gaede stated that all vapour pumps depended for their successful action on the fact that the gas enters the system through a diffusion diaphragm, defining the latter as an orifice through which gas and vapour were flowing in opposite senses and in which the mean free path of the gas molecules in the vapour was of the same order as the linear dimensions of the orifice. We have seen that this is true in a sense, since in all pumps the gas must enter the stream through a cloud of backstreaming vapour, and the mean free path of the gas molecules in this vapour must be large.

In this way all vapour pumps may be said to operate on the Gaede principle since they will only operate satisfactorily when backstreaming has been reduced until the mean free path of the incoming gas molecules in it is large. In the Gaede pump, backstreaming is reduced by adjusting conditions to give the correct stream density–slit-width relation, and in the Langmuir by directing the stream.

The effect of excessive heating

The critical effect of backstreaming on the entry of gas into the stream is confirmed by observations on the effect of increasing the temperature of the vapour stream without increasing its speed. In all pumps, as the vapour molecules get hotter and collisions more frequent, the amount of backstreaming increases and the speed of the pump falls off. Gaede has accounted for this in his formula for the speed of the pump, the coefficient α decreasing with increasing vapour pressure and temperature. In the Langmuir pump the speed falls off with increasing heat input (after a maximum has been passed), and if the walls are not sufficiently cooled to prevent emission from them, this will also cause a fall in the speed of pumping. Crawford finds that the speed of his pump falls off sharply with increasing heat input after a maximum value, but says that this is due to the density of the dispersing fringe exceeding a certain value. As Gaede points out, however, it is more likely that the fall in speed is due to an increased number of collisions in the stream resulting in increased backstreaming.

Absorptivity

In addition to the effect of backstreaming on the entry of gas into the stream, there is the secondary factor which Gaede calls "absorptivity". This factor takes into account the fact that the vapour stream is not capable of entraining and removing all the gas molecules which reach it. Gaede suggests that this is due to the saturating of the surface layers of the stream with gas, and he gives experimental evidence in confirmation of this, and suggests that in pumps of the Langmuir type, where the stream density increases rapidly with depth of penetration, saturation may occur more readily than in his own pumps, which utilize a stream of uniform density. Beyond this little is known of the effect, and none of the other authors quoted refer to it, though it seems highly probable that some such effect exists.

§ 5. DEVELOPMENTS SINCE 1923

The theoretical conclusions given above were developed from work done by Gaede, Langmuir and Crawford between 1915 and 1923 (*loc. cit.*), and there has been no further published work of a sufficiently fundamental nature since that date to affect the conclusions. Most of the more recent developments in the field have been in design, with the object of reducing the quantity of backstreaming vapour to a minimum. Of the papers published between 1923 and 1939, those of Ho (1932 a and b) require special mention in that he introduced the "Ho coefficient" for the efficiency of a vacuum pump. This coefficient is defined as the ratio of pump speed to the theoretical (Knudsen) admittance of the annular gap round the nozzle of a pump. Nearly all the pumps developed used the directed stream of the Langmuir pump and utilized nozzle forms to obtain high-velocity pumping streams as Crawford did. The first of these was that designed byGaede (1923), already referred to (figure 9 (*a*) and (*b*)), and which was subsequently improved (Gaede and Keesom, 1929).

The most recent developments are due to workers in America and are well summarized by Hickman in a recent paper (1940). The high-speed jets of Embree (U.S. Pat. 2,150,676) and Stallman (Brit. Pat. 565,455) are both designed to give a high-velocity downward-directed stream to reduce backstreaming and increase absorptivity, resulting in higher speeds for the reasons given above.

In a recent paper, Alexander (1946) described a successful high-vacuum pump, the design of which was based on the desirability of reducing the backstreaming to a minimum. In the discussion of the theory of the pump, however, several misconceptions arise which are commonly met with in such discussions. Alexander stated Gaede's condition for the operation of a diffusion pump as being that the mean free path of the vapour molecules in the stream shall be of the order of the width of the orifice. This statement, which is to be found in other books and papers on the subject, is incorrect.* The correct statement is that ***the mean free path of the gas molecules in the backstreaming vapour*** shall not be less than the dimensions of the orifice. In a Gaede type of diffusion pump (figure 2), for the fundamental Gaede condition to hold, that quoted by Alexander must hold as well, but, as has been shown, in diffusion pumps of the Langmuir or Crawford types, whilst in these pumps, too, the mean free path of the incoming gas in the backstreaming vapour must be of the order of the dimensions of the orifice, the pressure of the vapour in the stream can be considerably greater than that which would be satisfactory in a Gaede pump. From this it can be seen that the correct inference to be drawn from Alexander's calculations of the amount of backstreaming, which show that the density of the backstreaming vapour is very low, is that his pump does satisfy Gaede's principle, and is a " diffusion " pump. The fact that the ratio of the speed of the pump to the area of the orifice is not dependent on the throat width, as might be expected from a first consideration of Gaede's theory, is easily explained by saying that the jet of his pump reduces backstreaming so effectively that even at the largest throat-widths met with it is still insufficient materially to affect the speed of the pump.

A paper shortly to be published by Blears and Hill † deals with the performance of the diffusion pump with light gases.

§ 6. SUMMARY

For any diffusion pump to be successful it must satisfy two conditions :—

(1) The incoming gas must be able to enter the pumping stream easily.

(2) The pumping stream must effectively prevent the gas in the backing side from passing into the high vacuum side.

---

[*] This misconception probably arose because in Gaede's first paper (1915) he himself did not make it clear that the important factor was the mean free path of the gas molecules in the backstreaming vapour molecules. This was clarified in his 1923 paper, but for some reason the significance of this paper has been largely overlooked. In fact, the present authors were not familiar with it until recently and are very grateful to Mr. H. Griffiths of London for drawing their attention to the importance of Gaede's 1923 paper.

[†] Research Department, Metropolitan-Vickers Electrical Co. Ltd.

The first condition was stated by Gaede as: "the mean free path of the gas molecules entering the pump in the vapour molecules backstreaming from the pumping stream must not be less than the dimensions of the orifice". This condition may be simply stated as "backstreaming must be reduced to a minimum".

In the three major types of pump this is done in three different ways:

(i) In the Gaede-type pump (figure 2) backstreaming is reduced by increasing the mean free path of the vapour molecules in the stream. If this mean free path is increased until its value is greater than the linear dimensions of the orifice, then condition (2) breaks down. The pump operates most satisfactorily when the mean free path of the vapour molecules in the stream is of the order of the linear dimensions of the orifice.

(ii) In the Langmuir-type pump (figure 5), backstreaming is reduced by directing the stream in the pumping direction and condensing it on impact with the walls. It must be remembered, as a second-order effect, that the mean free path of the vapour molecules in the stream must not be too small a fraction of the dimension of the orifice.

(iii) In the Crawford-type pump (figure 7), backstreaming is reduced by directing the stream and giving the vapour molecules a very high velocity in the pumping direction.

In any pump, increasing the velocity of the vapour stream tends to increase the effective mean free path of the vapour molecules in the stream, and so further reduce backstreaming.

Most modern diffusion pumps are a combination of the Langmuir and Crawford types.

§ 7. EXPERIMENTAL CHARACTERISTICS

It can now be shown how the working characteristics of a simple form of modern diffusion pump may be explained qualitatively by means of the theory developed above. The characteristics considered are those of a small diffusion pump, a "Metrovac" type 03 (figure 11) having a single umbrella-type jet and operating with "Apiezon B" oil. This form of jet is typical of those in common use today.

It must be remembered that, whilst the general form of the characteristics given is the same for all pumps and may be correlated with the theory, any quantitative values quoted are critically dependent on the particular size of pump and type of working fluid, and cannot be used as checks for the theory without a more precise knowledge of these factors.

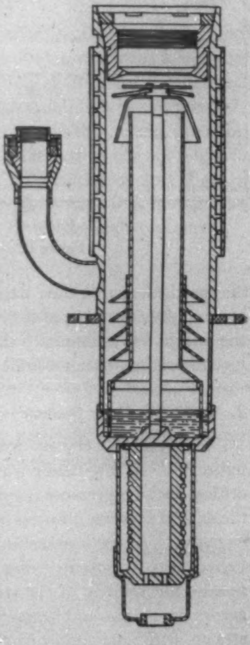


Figure 11.

1. *The relation between speed and heater wattage*

The curve obtained by plotting values of the speed, at a very low backing pressure ($<1\,\mu$), against the heater wattage is shown in figure 12. The general form of this characteristic shows an optimum value of heater wattage, the maximum in the speed values not being sharply defined. The explanation follows directly from the theory. At low heater wattages, where the vapour pressure is just less than or equal to the backing pressure employed, the stream is insufficient to seal the pump, and thus it has zero speed. As the heater wattage is increased the sealing becomes more efficient. This improvement is offset by an increase in backstreaming; this follows since the increased temperature of the stream causes an increased number of collisions between the vapour molecules of the stream. For a given backing pressure, once the pump acts as an efficient seal, an increase in the heater wattage merely serves to increase the backstreaming and hence decrease the speed, the curve thus passing through a maximum. Obviously, for a greater backing pressure (dotted curve in figure 12), more heater watts are required to seal the pump. Consequently the optimum value of the speed is less than with the lower backing pressure, since it occurs at a higher vapour temperature.

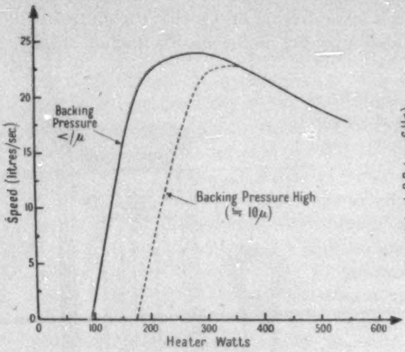


Figure 12.

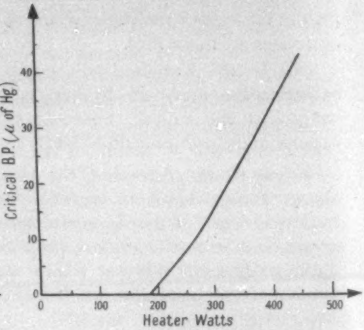


Figure 13.

2. *Critical backing pressure—heater-watts relation*

The relation between critical backing pressure and heater watts is shown in figure 13. At low-heater wattages, where the vapour pressure is very small, the critical backing pressure is practically zero. As the heater watts are increased, so the critical backing pressure increases, the relation becoming nearly linear at higher wattages. As the wattage increases, the temperature of the stream increases, and hence, also, its potential energy. Again, for a given throat-width in the pump nozzle, the velocity of the stream increases, due to a greater pressure difference across the nozzle, and hence the kinetic energy of the stream also increases. The stream, with total energy increased, is thus able to seal off a larger pressure on the backing side.

3. *The relation between fine side pressure and backing pressure and between speed and backing pressure*

As has been discussed, the amount of gas back-diffusing through the stream, for heavy gases, is practically zero for backing pressures below the critical, and above the critical this amount increases very rapidly. This satisfactorily explains the shape of the experimental curve shown in figure 14.

Since the speed of the pump may be regarded as the net rate of transfer of gas in the pumping direction—that is, the total amount passed in the pumping direction minus that diffusing back—the shape of the curve of speed against backing pressure may easily be derived from the curve above. An experimental curve is shown in figure 14, the speed remaining independent of the backing pressure until the critical value is passed, when the speed falls off sharply.

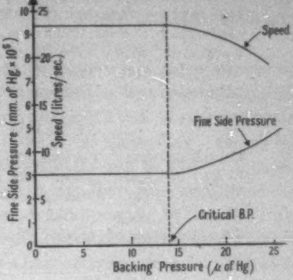


Figure 14.

4. *The relation between speed and fine side pressure*

We have seen that, in the absence of back-diffusion of gas, the pumps discussed have no limiting vacuum, and hence, under these conditions, the speed is independent of the fine-side gas pressure. If the back diffusion is not zero, then the speed (regarded as the net transfer of gas in the pumping direction) falls to zero when the transfer of gas in the pumping direction is equal to that in the reverse direction. In this condition the pump merely passes gas from the fine to the backing side, whence it returns by diffusing back to the fine side. When the amount of gas on the fine side due to back diffusion is negligible compared with that due to other sources, then the speed of the pump is independent of the fine-side pressure.

§ 8. CONCLUSION

It has been shown that the modern diffusion pump is an outcome of the original pioneer work of Gaede, Langmuir and Crawford. When observed in the true perspective, it is seen that the contributions made by these three were not competitive, but rather were complementary, the best pumps depending on advances made by all three.

We have not tried to apportion the credit for any particular part of the theory, but rather to synthesize a complete theory based on all the data available.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Mr. T. M. Hodson for the translation of Gaede's papers and also to Sir Arthur P. M. Fleming, C.B.E., D.Eng., Director of Research and Education, and Mr. B. G. Churcher, M.Sc., M.I.E.E., Manager of Research Dept., Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

BIBLIOGRAPHY

Alexander, 1946. *J. Sci. Instrum.*, **23,** 11.
Crawford, 1917. *Phys. Rev.*, **10,** 557.
Gaede, 1915. *Ann. der Phys.*, **46,** 357.
Gaede, 1923. *Z. tech. Phys.*, **4,** 348.
Gaede and Keesom, 1929. *Zeit. f. Instrum.*, **49,** 298.
Hickman, 1940. *J. Appl. Phys.*, **11,** 303.
Ho, 1932 a. *Rev. Sci. Instrum.*, **3,** 133.
Ho, 1932 b. *Amer. Phys. Soc. Proc.*, Jan.-June 1932, 386.
Langmuir, 1916. *Gen. Elect. Rev.*, **19,** 1060.
Molthan, 1925. *Phys. Z.*, **26,** 712.
Molthan, 1926. *Z. Phys.*, **39,** 1.

GENERAL BIBLIOGRAPHY

Dunoyer, 1924. *Vacuum Practice.*
Goetz, A., 1926. *Physik und Technik des Hochvakuums* (Chaps. 1 and 2).
Newman, 1925. *Production and Measurement of Low Pressure* (Chap. 5).
Witty, R., 1945. *J. Sci. Instrum.*, **22,** 201.
The Dictionary of Applied Physics.

THE EFFECT OF TEMPERATURE ON THE THERMAL CONDUCTIVITY AND VISCOSITY OF LIQUIDS

By R. EISENSCHITZ,
University College, London

Communicated by E. N. da C. Andrade, F.R.S. ; *MS. received* 26 *August* 1947

ABSTRACT. The theory of Brownian movement is applied to the ‘cell model’ of liquids. The Smoluchowski equation for the density distribution is solved under conditions corresponding to laminar flow and conditions corresponding to inhomogeneous temperature. The flow of momentum and the flow of energy are calculated. The result indicates that the coefficient of viscosity varies exponentially with the inverse temperature and that the coefficient of thermal conductivity is proportional to the absolute temperature.

§ 1. INTRODUCTION

As early as 1885 Graetz noticed that the effect of temperature on the viscosity and thermal conductivity of liquids is of opposite sign and different magnitude, whereas in gases these two transport coefficients are known to be almost proportional to each other. Graetz argued that the intermolecular forces have only an indirect influence on the mechanism of heat transfer, but exercise resistance against deformation.

The relation between temperature and viscosity has subsequently been found to follow an exponential law, the significance of which for the kinetic theory of liquids was first recognized by Andrade (1934). Little attention has, however, been given to the differences between the two transport phenomena, which are so marked that even a very rough model of the liquid state should indicate their origin.

The model of a liquid which is used in this paper is chosen from the point of view of mathematical simplicity rather than the close expression of experimental data. A model of this kind is the 'cell model', representing a molecule in the field of force of its neighbours. The motion of the representative molecule may be regarded as Brownian movement and, in particular, the distribution in space be supposed to be governed by the Smoluchowski equation.

Although this is a crude model, it is in accordance with Kirkwood's theory (1946) of irreversible processes, in which he succeeds in deriving irreversible changes of the average state of motion of one molecule or of a pair of representative molecules from the reversible mechanics of the totality of molecules. He shows that the distribution of a pair of molecules in coordinates and momenta is determined by an equation of the type that applies to the Brownian movement of colloid particles, and he derives the friction constant appearing in this equation from molecular data. His theory of viscosity and thermal conductivity requires the determination of the distribution in phase of a pair of molecules which depends on the solution of a partial differential equation with twelve independent variables.

The comprehensive and systematic kinetic theory of liquids being worked out by Born and Green (1946, 47, Green, 1947) has already given valuable results, but by the nature of the rigorous methods employed is not likely to lead to simple formulae.

By considering the distribution of one representative molecule instead of a pair—that is by means of the cell method—fairly good results are obtained in the field of equilibrium theory. The differential equation in the theory of Brownian movement determining the distribution in momenta and coordinates may be replaced by the Smoluchowski equation in coordinates only, if the friction constant is sufficiently high.

A method similar to that used in this paper has been employed by Kramers (1940) for calculating the rate of transitions of particles over a potential barrier.

The cell model will accordingly be used for calculating the transport coefficients of the liquid in terms of the friction constant and of the particular law of force within the cell. The latter is assumed to be proportional to the distance from the centre, i.e. to be derived from a parabolic potential field increasing from the centre to the surface of the cell and remaining at constant value outside the surface.

§ 2. THE SMOLUCHOWSKI EQUATION

Consider a spherical cell of radius R, volume V_0, in which the potential energy is a function of the distance from the centre. The mean number density of the liquid is accordingly $(N/V)=(1/V_0)$. The probability distribution of the molecule in the cell, to be denoted by $g(r, \theta, \phi)$ is in thermal equilibrium, given by

$$g=g_0=(1/Z).\exp(-U/\boldsymbol{k}T)$$

$$Z=\int \exp(-U/\boldsymbol{k}T)r^2 \sin\theta \, dr \, d\theta \, d\phi.$$

A non-equilibrium distribution which is independent of the time may be written

$$g = g_0(1+w),$$

where the function w is determined by the Smoluchowski equation (Chandrasekhar, 1943)

$$\operatorname{div}(\exp(-U/\boldsymbol{k}T) \,.\, \operatorname{grad} w) = 0, \qquad \ldots\ldots(1)$$

and is related to the flow vector according to

$$\mathbf{j} = (-\boldsymbol{k}T/\beta m) \,.\, \operatorname{grad} w. \qquad \ldots\ldots(1\,a)$$

In this equation m is the mass of the molecule and β is the "friction constant" of dimension $(\text{time})^{-1}$.

In the present theory equation (1) is assumed to hold in the interior of the cell even when the temperature or macroscopic velocity of the liquid is inhomogeneous. The effects of the departure from equilibrium are accounted for by conditions which the function w has to satisfy on the surface of the cell and according to which the density and the normal component of flow are made to be continuous on the surface. It should be said at once that these conditions can be satisfied only by functions with a singularity in the cell.

In the case of laminar flow in the direction of the y axis and of the rate of shear $2a$, the postulate of continuity of the normal component of flow requires that for $r = R$

$$(\boldsymbol{k}T/m) \,.\, g_0 \,.\, (\partial w/\partial r) = (ar/2NV_0) \,.\, \sin^2\theta \,.\, \sin 2\phi.$$

The factor $1/NV_0$ on the right-hand side accounts for the magnitude of the mean flow of one molecule. This condition is satisfied by putting

$$w = u(r) \,.\, \sin^2\theta \,.\, \sin 2\phi \qquad \ldots\ldots(2)$$

and demanding that

$$du/dr = [(armZ/2N\boldsymbol{k}TV_0) \,.\, \exp(U/\boldsymbol{k}T)], \qquad (r=R). \qquad \ldots\ldots(3\,a)$$

Continuity of density is possible only if

$$u = 0, \qquad (r=R). \qquad \ldots\ldots(3\,b)$$

If there is a gradient of temperature

$$T = T_0(1 + br \,.\, \cos\theta)$$

the density at the surface is determined by

$$g_0(1+w) = (1/NV_0) \,.\, \exp(-U(1 - br \,.\, \cos\theta)/\boldsymbol{k}T_0)$$

or, writing T for T_0,

$$g_0 \,.\, w = br \,.\, \cos\theta \,.\, (U/\boldsymbol{k}T)(1/NV_0) \,.\, \exp(-U/\boldsymbol{k}T),$$

which is satisfied by putting

$$w = v(r) \,.\, \cos\theta \qquad \ldots\ldots(4)$$

and demanding that

$$v = brZU/NV_0\boldsymbol{k}T, \qquad (r=R). \qquad \ldots\ldots(5\,a)$$

When the flow of heat is steady there is no flow in the liquid, so that the normal component of flow must vanish at the surface of the cell:

$$dv/dr = 0, \qquad (r=R). \qquad \ldots\ldots(5\,b)$$

It will be seen that these surface conditions are decisive for the effect of temperature on the transport coefficients.

§ 2. SOLUTION OF THE EQUATION

The potential energy is, in the following, assumed to be

$$U=\tfrac{1}{2}Gr^2,$$

so that the model requires altogether four constants for its complete specification, the mass m, the friction constant β, the force constant G and the cell radius R.

Introducing a dimensionless variable,

$$\xi=(G/2\mathbf{k}T)^{\frac{1}{2}}\,.\,r,$$

the equilibrium distribution is

$$g_0=(1/Z)\,.\exp(-\xi^2), \qquad Z=(2\pi\mathbf{k}T/G)^{3/2}.$$

In order to find a solution of equation (1) which is appropriate to laminar flow, equation (2) is introduced into (1). From this a differential equation of u as function of ξ is obtained:

$$\xi^2u''+2\xi(1-\xi^2)u'-6u=0, \qquad \text{......(6)}$$

and solved by standard methods. The general solution is

$$u=A_1u_1+A_2u_2,$$

where A_1 and A_2 are constants and

$$\left.\begin{aligned} u_1&=\left[(1/\xi^3)\,.\exp\xi^2\,.\int_0^\xi \exp(-t^2)\,dt\right]-(1/\xi^2)-(2/3),\\ u_2&=-\left[(1/\xi^3)\,.\exp\xi^2\,.\int_\xi^\infty \exp(-t^2)\,dt\right]-(1/\xi^2)-(2/3).\end{aligned}\right\} \qquad \text{......(7)}$$

Similarly a solution of equation (1) is found which is appropriate to inhomogeneous temperature. Equation (4) is introduced into (1) so that a differential equation for v is obtained:

$$\xi^2v''+2\xi(1-\xi^2)v'-2v=0. \qquad \text{......(8)}$$

It has the general solution

$$v=B_1v_1+B_2v_2,$$

where B_1 and B_2 are constants and

$$\left.\begin{aligned} v_1&=\left[(2+1/\xi^2)\,.\int_0^\xi \exp t^2\,dt\right]-[(1/\xi)\exp\xi^2],\\ v_2&=2+(1/\xi^2).\end{aligned}\right\} \qquad \text{......(9)}$$

The functions u_1, v_1 increase exponentially for large values of ξ. The functions u_2, v_2 are finite at infinity and have a pole at $\xi=0$. Since the density of force remains finite and the integral over w vanishes, there is no necessity for rejecting this calculation in spite of the singularities.

In the following calculations the expressions for the functions u and v are simplified in order to avoid lengthy formulae. The following first terms of expansions are used:

for $\xi \gg 1$

$$\begin{aligned} u_1&=\tfrac{1}{2}\sqrt{\pi}\,.(1/\xi^3)\exp\xi^2 & u_1'&=\sqrt{\pi}\,.(1/\xi^2)\exp\xi^2\\ u_2&=-2/3 & u_2'&=2/\xi^3\\ v_1&=(1/\xi^3)\exp\xi^2 & v_1'&=(2/\xi^2)\exp\xi^2;\end{aligned}$$

for $\xi \ll 1$

$$u_2 = -\tfrac{1}{2}\sqrt{\pi}\,.\,[(1/\xi^3)+(1/\xi)+(\xi/2)].$$

The constants of integration are determined by inserting the expressions for u and v into equations (3) and (5). In this calculation the expansions for large values of ξ are used. The result is

$$\left.\begin{aligned} A_1 &= 3\beta ma/4NG, \\ A_2 &= \frac{(9\sqrt{(2\pi)}\beta ma)\,.\,(\boldsymbol{k}T)^{3/2}\,.\,\exp(GR^2/2\boldsymbol{k}T)}{4NG^{5/2}R^3}; \end{aligned}\right\} \quad \ldots\ldots(10)$$

$$\left.\begin{aligned} B_1 &= \frac{(3\sqrt{\pi b\boldsymbol{k}T})\,.\,\exp(-GR^2/2\boldsymbol{k}T)}{4NRG}, \\ B_2 &= 3\sqrt{\pi}(\boldsymbol{k}T)^{1/2}b/2^{3/2}NG^{1/2}. \end{aligned}\right\} \quad \ldots\ldots(11)$$

The constants A_1, B_1 are clearly of smaller magnitude than the constants A_2, B_2. The constant A_2 contains an exponential factor $\exp(GR^2/2\boldsymbol{k}T)$ which appears in the coefficient of viscosity. The constant B_2 has no exponential factor.

§4. THE TRANSPORT COEFFICIENTS

The distribution functions u, v are used for calculating the flow of momentum and the flow of energy. In a dilute gas these quantities are determined by the amount of momentum and kinetic energy which isolated molecules carry along their path. In dense gases the transfer of momentum and energy during collisions is appreciable (Chapman and Cowling, 1939); the same mechanism of transfer prevails in liquids and solids, although there are no "collisions". Momentum and energy are continually exchanged between all those molecules for which the force of interaction is appreciable. Roughly, the flow of momentum is given by the negative product of force and distance, the flow of energy by the product of potential energy and relative velocity plus product of force, relative velocity and distance.

The flow of momentum and energy can be calculated according to formulae given by Kirkwood (1946). These formulae depend upon the distribution of two molecules in coordinates and moments, but can be rewritten in such a way that they apply to the cell model for which the distribution in coordinates only is given. This is obvious for the flow of momentum, which is equal to the stress produced by intermolecular forces. The flow of energy depends explicitly upon the momentum, but only in such a way as to be proportional to the flow of one molecule relative to another molecule. This flow can be derived from the distribution in relative coordinates, according to equation (1 *a*).

Denoting the orthogonal components of the position vector by s_i and of the flow vector by j_k $(i, k=1, 2, 3)$, the components of the stress tensor are

$$P_{ik} = (N/2V_0)\,.\,\overline{[(\mathrm{d}U/\mathrm{d}r)(s_i\,.\,s_k)(1/r)]}$$

and the density of energy flow is

$$Q_i = (N/2V_0)\,.\,\overline{(Uj_i - [\tfrac{1}{2}(\mathrm{d}U/\mathrm{d}r)\sum_k(s_i j_k + s_k j_i)s_k])},$$

where the bar indicates the average over the distribution function.

The shearing stress is accordingly

$$P_{xy}=(N/2V_0)\iiint g_0\,.\,(\mathrm{d}U/\mathrm{d}r)ur^3\sin^5\vartheta\sin^2 2\phi\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

The components of energy flow are

$$Q_r=(N/2V_0)\iiint[U-(r\,\mathrm{d}U/\mathrm{d}r)]j_r r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi,$$

$$Q_\theta=(N/2V_0)\iiint[U-\tfrac{1}{2}(r\,\mathrm{d}U/\mathrm{d}r)]j_\theta r^2\sin\theta\,\mathrm{d}r\,d\theta\,\mathrm{d}\phi,$$

$$Q_z=Q_r\cos\theta-Q_\theta\sin\theta.$$

The components of flow are

$$j_r=-g_0(\mathbf{k}T/\beta m)(\mathrm{d}v/\mathrm{d}r)\cos\theta,$$
$$j_\theta=g_0(\mathbf{k}T/\beta m)(v/r)\sin\theta,$$

so that

$$Q_z=(N\mathbf{k}T/2\beta mV_0\iiint(g_0[(r\,\mathrm{d}U/\mathrm{d}r)-U)(\mathrm{d}v/\mathrm{d}r)\cos^2\theta$$
$$+(\tfrac{1}{2}r\,(\mathrm{d}U/\mathrm{d}r)-U)(v/r)\sin^2\theta]r^2\,.\,\sin\theta\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

In calculating P_{xy} and Q_z the terms in A_1 and B_1 are neglected, the function u_2 is represented by its expansion for low values of ξ, and integration over the radius is extended to infinity. The result may be formulated in terms of the coefficient of viscosity η and the coefficient of thermal conductivity λ:

$$P_{xy}=-2a\eta=-2a\,(27/40\sqrt{(2\pi)})[m\beta(\mathbf{k}T)^{5/2}/R^6G^{5/2}]\exp(GR^2/2\mathbf{k}T), \quad\ldots\ldots(12)$$

$$Q_z=-bT\lambda=-bT\,(3/8\pi)(\mathbf{k}^2T/m\beta R^3). \quad\ldots\ldots(13)$$

§ 5. CONCLUSION

The coefficients of viscosity and thermal conductivity are thus obtained in terms of the friction constant. If this is supposed to depend only slightly upon the temperature, the two formulae give a reasonable representation of the type of temperature dependence of these two coefficients. Whereas the model and particularly the parabolic potential-energy curve cannot be expected to provide quantitative agreement with experiment, it is important to note that the reason for the different temperature effects is to some extent independent of the special assumptions. The only significant difference between the calculation of viscosity and thermal conductivity is in the boundary conditions. Their physical significance can be said to be the fact that molecules are forced to travel over regions of high potential energy in viscous flow, but not in a gradient of temperature. In the latter case they only carry out a circulation within the cell which distinguishes the case of a temperature gradient from that of equilibrium conditions.

This difference in mechanism between thermal conduction and viscous flow is independent of the present model and likely to be the explanation for the observed behaviour.

ACKNOWLEDGMENT

The writer is indebted to Professor E. N. da C. Andrade, F.R.S., for his interest in this work and for valuable discussions.

REFERENCES

Andrade, 1934. *Phil. Mag.*, **17**, 497, 698.
Born and Green, 1947. *Nature, Lond.*, **159**, 251 ; 1946. *Proc. Roy. Soc.*, A, **188**, 10.
Chandrasekhar, 1943. *Rev. Mod. Phys.*, **15**, 1.
Chapman and Cowling, 1939. *The mathematical theory of inhomogeneous gases* (Oxford).
Graetz, 1885. *Ann. d. Phys., Lpz.*, **18**, 336.
Green, 1947. *Proc. Roy. Soc.*, A, **189**, 103.
Kirkwood, 1946. *J. Chem. Phys.*, **14**, 180.
Kramers, 1940. *Physica*, **7**, 284.

DISCUSSION

on paper by Dr. R. E. Siday, entitled "The Optical Properties of Axially Symmetric Magnetic Prisms" (*Proc. Phys. Soc.*, **59**, 905 (1947)).

Dr. O. Klemperer. Everybody who has worked practically with deflecting fields in velocity analysis or in spectrographs will have felt the great need which Dr. Siday has satisfied by attacking the problem of the magnetic prism from the theoretical side. The ballistic or electrodynamical approach is very complicated, and it implies great advantages to be able to treat the deflecting fields in terms of the optical methods. When I tried, many years ago, to apply the principles of design of the optical prism-spectrograph to β-ray spectroscopy, I was surprised to find the lines of β-spectra at all. However, these lines, or rather patches, were so broad that the lack of definition almost upset the high theoretical resolving power of the new instrument. Dr. Siday's method of cancelling the positive spherical aberration of the lenses by a negative aberration of the prism seems very promising. I should like to ask how far it will be possible to match the cylinder optics of the prism with the spherical optics (circular symmetry) of the lenses ? Does Dr. Siday expect that the lines of his spectra will be really straight or very much curved ?

Dr. H. H. Hopkins. It will be interesting to see the results of Dr. Siday's investigation of rays not in the plane of symmetry, but, irrespective of what he finds for these rays, the prism that he has devised should still prove very useful. If the optical analogy holds, one could use a rectangular aperture with its longer dimension parallel to the plane of symmetry to permit only the corrected rays to form the image, and this without serious loss of resolving power in the direction perpendicular to the line-images.

Author's reply. Both Dr. Klemperer and Dr. Hopkins raise points of considerable importance, which will be considered in detail in Part II of this paper. In anticipation I would reply to Dr. Klemperer that by limiting the beam very straight lines can be obtained, and to Dr. Hopkins that I propose to use the prism in the way he suggests and I agree that over a limited aperture the resolution should not be seriously reduced.

DISCUSSION

on papers by C. Gurney entitled "Delayed Fracture in Glass" (*Proc. Phys. Soc.*, **59**, 167 (1947)), and R. A. Sack entitled "Extension of Griffith's Theory of Rupture to Three Dimensions" (*Proc. Phys. Soc.*, **58**, 729 (1946)).

Prof. G. I. Finch referred to the experimentally established existence of the cracks. He attributes their failure under load to adsorption of gases (H_2O, O_2, CO_2 in particular) towards the edge (end) of the crack exerting a wedging action. He drew attention to the adsorption of such gases on the fresh mica cleavage surface and the effect which such adsorption has on the re-packing of cleavage faces.

Mr. W. C. Hynd. Can the author say if cracks have been observed by the etching method adopted by Andrade on freshly blown samples of glass as distinct from mechanically polished surfaces ?

Mr. J. H. Awbery. The general idea of Griffith's argument is so appealing that one would be surprised if it did not agree with the closer analysis presented by atomic theory. Are they really independent, however ? The Griffith theory gives its criterion in terms of E and γ, but does not predict their values. When the observed values are inserted, are we not inserting the results which would follow from a successful atomic analysis of the phenomena ?

Mr. S. M. Cox. To subscribers to the Griffith flaw theory, Mr. Gurney's paper must supply a great deal of food for thought, but to a non-subscriber it seems rather to illustrate the expedients to which one is driven in order to account for the known facts on the basis of a theory which, in my view, begs the question.

I hope to show in a forthcoming paper that if one treats the problem from the beginning—that is to say, before the formation of the flaw instead of after it—and assumes that, under the conditions of test, the formation of such a flaw consitutes the trigger action for the rupture, then, at least for glass, all the salient facts are readily explained. The fact that rupture always progresses with a finite velocity I take to indicate that the theoretical strength (so called) is never actually reached.

The outstanding points are :

(1) The comparative weakness of the surface.

(2) The considerable spread of results under apparently identical testing conditions.

(3) The influence of time in producing eventual breakage without apparently weakening the glass before the breakage occurs.

(4) The apparent increased strength of fine fibres.

Taking into consideration the thermal energy of the ions together with the contribution due to the applied stress, the probability of the formation of a flaw of w ions may be shown to be proportional to

$$T\left\{N\left(e^{\frac{-B-S}{k\theta}}-2e^{\frac{-B+S}{k\theta}}\right)\right\}^{w},$$

where B is the potential barrier energy and S is the contribution to the total vibrational energy made by the applied tension. T is the duration of the test and N is the total number of ions.

If this expression does in fact define the condition for rupture, then it follows at once that :

(*a*) The surface is much weaker than the bulk except where the sample is exceedingly large, because the number of ions required to form a surface flaw is only about half that required for a non-surface flaw.

(*b*) The spread of the experimental results using one fiducial constant (viz. the mean strength) is given more closely than by a Gaussian distribution using two.

(*c*) There is excellent agreement with Preston's " Duration of Test " experiments.

(*d*) The strength of fine fibres should increase with diminishing radius as shown by Griffith's results but the agreement is only good down to about 0·05 mm. radius.

Where surface flaws, atmospheric attack, etc., are bound to play their part in practice, I believe the underlying explanation of the test results lies in the formation of flaws and not in their pre-existence.

Mr. R. W. Douglas. Experiment shows that the surface of glass has a very profound effect on its strength, particularly if it contains macroscopic cracks. Macroscopic cracks are easily formed on glass surfaces by subjecting the surface to some treatment, at a temperature a hundred degrees or so above room temperature, which removes alkali from the surface, e.g. attack by wet SO_2. This reduces the thermal expansion of the surface layers and on cooling the stresses set up by the differential contraction result in a system of very fine cracks. It is possible that a similar process in the ordinary manufacture of glass might produce such cracks on a microscopic scale. It is improbable therefore that effects such as

Mr. Cox has discussed can account for all the phenomena exhibited in the brittle rupture of glass. However, I agree with him in doubting that the Griffith crack mechanism is the only operative one.

X-ray diffraction experiments show that the arrangements of the atoms in a glass are exactly similar to those in a liquid if one could imagine all the translational motion in the liquid stopped. Now Temperley (*Proc. Phys. Soc.*, **58**, 436, 1946) has found that the tensile strength of water is much less than the strength predicted from theory. It is difficult to see how the Griffith theory can be applied to water and one wonders whether in glasses and in water there is not some other mechanism controlling rupture, and in view of the random atomic arrangements, whether it is possible that the estimated theoretical values of the strength are too high.

Mr. H. A. ELLIOTT. Dr. Sack suggested that the normal Griffith criterion appeared to differ from that given by the condition that the stress at the apex of the crack should be equal to the theoretical strength of the material. I have shown (Elliott, 1947), however, that the theoretical strength condition leads to the relation $T \simeq P\sqrt{a/c}$, where P=theoretical strength, T=applied stress, c=semi-length of crack and a=lattice spacing. Also the surface energy, $S \simeq P^2a/E$ where E is Young's modulus (this would hold on dimensiona grounds for a true law of atomic force as well as for the approximate one used in the quantitative calculation). Thus we see that $T \simeq \sqrt{ES/c}$ as Griffith finds (the numerical factor is relatively unimportant as it is of order unity) ; in the case of Griffith's experiments agreement is found numerically to approximately 10% error.

I agree with Gurney's statement that it seems unlikely that evaporation at the bottom of a crack is the cause of the slow fracture of glass under small loads. It should be observed that such a mechanism gives a law $t \simeq A\left(1-e^{-\lambda\frac{T^2}{P^2}}\right)$, which does not agree even in form with the result of Preston (1946). It may, however, be possible to find a mechanism due to chemical attack (with water vapour) at the apex, the rate being mainly controlled by a diffusion process through the gel or alkali salts formed after attack. So far as I am aware this problem has not been solved. Dr. Orowan (1944) suggests a similar mechanism, quoting the fracture of mica. He also points out that the Griffith formula gives the correct strengths for both rapid and slow fracture if the vacuum value of the surface energy is used in the first case and the air-surface value in the second.

AUTHOR'S reply. In reply to Mr. Hynd, Andrade (1937) found very few cracks on etching freshly blown glass, but they appeared on etching glass which had been exposed to air for some hours after blowing. This result is consistent with the origin of the cracks suggested in the paper, but no doubt there are other possible explanations.

In reply to Mr. Awbery, although the Griffith theory does not predict the values of E and γ, the values of these constants are not chosen so as to make the predictions of the Griffith theory agree most closely with experiment. For low stresses E is readily obtained by direct measurement, as is γ at high temperatures. The values inserted in the Griffith formula should be the values predicted by extrapolation of these direct measurements. The calculation of E and γ depends on quantum mechanics whereas the Griffith theory uses only classical ideas. In a more complete theory therefore it seems likely that the division of the subject into the prediction of E and γ and the utilization of these and other macroscopic quantities in a theory of fracture would still be convenient.

With regard to the remarks of Mr. Cox, I had thought that the activation energy necessary for the spontaneous formation of flaws was too high compared with the thermal energy at room temperature for the process he describes to be an appreciable factor in the strength of glass. If such an effect were predominant in causing delayed fracture it would seem that the theoretical strength of glass rods might be approached at very rapid rates of loading—a prediction which could perhaps be tested by experiment.

In reply to Mr. Douglas, a more recent paper by Temperley (1947) reduces the ratio of experimental to theoretical tensile strength of water to about 1 : 10. This ratio would probably be further reduced if Temperley's experiments could be repeated using more rapid rates of stressing. If an important delayed fracture effect were found for water a theory on the lines of that outlined by Mr. Cox might be applicable.

With regard to Mr. Elliot's remarks, I agree that the form of my equation (13) does not give a very good fit to Preston's results, and that the rate of diffusion of atmospheric constituents through the complex of glass and atmospheric constituents formed at the base of the crack may be an important factor in determining the time to fracture. The argument I put forward for the non-contamination of the material at the end of the crack during evaporation (p. 178 of the printed paper) would not apply to processes of chemical attack.

In conclusion I would like to emphasize that, although Griffith's theory may give results of the right order of magnitude, it ignores the time taken for the crack to spread. In solids there is very appreciable delay in the attainment of states of lower free energy, often so much so, that states having free energy above the minimum possible persist indefinitely. It is no more correct to assume that failure occurs immediately free energy considerations are favourable than it is to assume that glass will crystallize or evaporate in the same circumstances.

REFERENCES

Andrade, A. N. da C., 1937. *Proc. Roy. Soc.*, A, **159**, 346.
Elliott, H. A., 1947. *Proc. Phys. Soc.*, **59**, 208.
Orowan, 1944. *Nature, Lond.*, **154**, 341.
Preston, 1946. *J. Appl. Phys.* (March).
Temperley, H. N. V., 1947. *Proc. Phys. Soc.*, **59**, 199.

DISCUSSION

on paper by J. C. Evans entitled "The Determination of Thermal Lagging Times" (*Proc. Phys. Soc.*, **59**, 242 (1947)).

Mr. D. K. Ashpole. Although Dr. Evans' paper deals primarily with the thermal lagging times of mercury barometers, reference is made to other possible applications. It therefore seems worth pointing out that care is necessary in applying the formulae to cases in which convection is important.

In the above paper it is assumed that the heat reaching any point in a column of liquid enclosed in a hollow cylinder is conducted either through the cylinder or through the liquid. This is only strictly true for an infinitely viscous liquid (i.e. a solid) due to the finite convection currents caused by the temperature gradient. Convection, however, will be negligibly small in comparison with conduction if the thermal conductivity of the liquid is high and the bore of the cylinder small. This accounts for the good agreement obtained between theory and practice for the mercury column defined in § 3 B.

For liquids other than mercury, the discrepancy between the theoretical and observed lagging times may become considerable. A case in point is a glass cylinder of 8·65 cm. bore, 0·35 cm. wall thickness, and 10 cm. long, containing water. A value of λ of 200 min. is obtained from Evans' equation (3), p. 246, namely,

$$\lambda_r = \frac{\rho_g S_g}{4k_g}(r_2^2 - r_1^2) + \frac{\rho_m S_m - \rho_g S_g}{2k_g} r_1^2 \log_e \frac{r_2}{r_1} + \frac{\rho_m S_m}{4k_m}(r_1^2 - r^2),$$

where ρ, S and k are the density, specific heat and thermal conductivity respectively, and the suffix g denotes the wall material, and the suffix m the liquid medium. The value $r=0$ was taken since the temperature was measured experimentally by means of a thermometer at the axis of the cylinder. The value of the lagging time obtained experimentally in a similar manner to that described by Evans was 9 min. In this instance, therefore, the convection is of greater moment than the conduction. Indeed, a better estimate of the lagging time is obtained by assuming perfect stirring within the cylinder (which gives a lower limit for λ) when a calculated value of 1 min. is obtained.

The applications of the equations derived by Evans should therefore be limited to cases where the effect of convection in the liquid can be shown or safely assumed to be negligible in comparison with conduction through it.

In the light of this conclusion, the value given on p. 248 for λ_0 with toluene in a thin-walled glass tube needs correction.

Author's reply. The theory given in the paper applies only to those cases in which convection effects are negligible, so that the results it gives will be erroneous when this condition is violated. Mr. Ashpole's note expresses the condition clearly.

OBITUARY NOTICES

FRIEDERICH PASCHEN

By a communication from his daughter, we have learned with deep regret that Professor F. Paschen died in Potsdam on 26 February 1947, at the age of 82. He was probably the greatest experimental spectroscopist of his time, and his record of discoveries is imposing both in its variety and in its importance. Fifty-five years have passed since Paschen's first classical publications on infra-red spectra appeared, and as the years progressed, his contributions to spectroscopy became more and more fundamental.

His earliest important publication on infra-red spectroscopy appeared in 1892. In the following year the extremely sensitive Paschen galvanometer was invented, and using it with a bolometer, Paschen was able to produce a deflection of 1 mm. for a temperature change of only 10^{-6} °c., a considerable achievement at the time. In 1894 he published the now classical determinations of infra-red refractive indices and dispersions for fluorite, rock-salt and quartz, and his values are still quoted in tables. This field was to interest him for several years to come, and although most of his energies were devoted to infra-red studies, yet it is of interest to note that as far back as 1895 he made measurements on the visible series spectrum of the newly discovered helium.

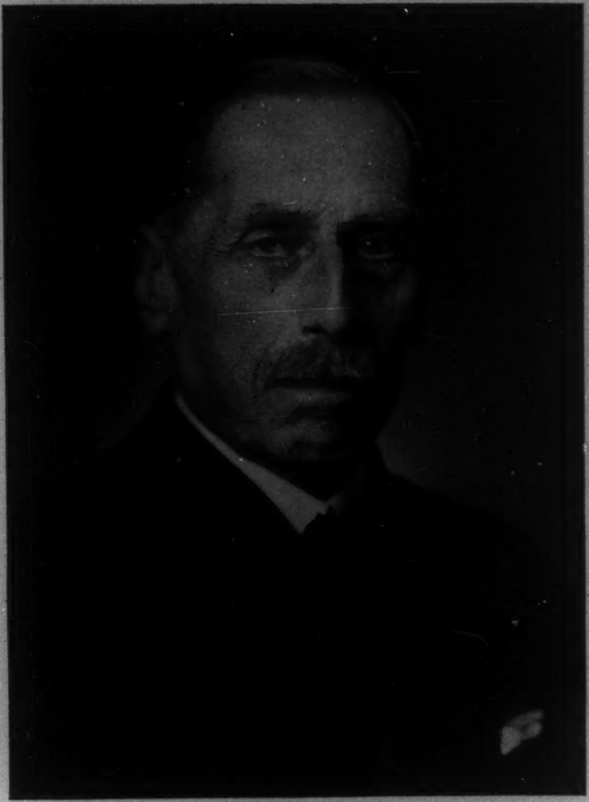
In 1899 Paschen founded the precision study of spectral black body radiation, and it is not widely enough realized that his pioneering work (which was followed by that of Lummer and Pringsheim) laid the experimental basis for the checking of Planck's quantum radiation formula. It was in this research that Paschen first introduced the use of concave mirrors for infra-red spectrographs, thus avoiding the considerable chromatic difficulties then inherent in the use of lenses. This technique is, of course, now absorbed into spectrography.

Paschen was always in the front line of progress, and in 1902 switched his attention to the investigation of the Zeeman effect, his studies of which were later to prove very fruitful indeed. His activities now ranged widely. In 1908 he found the infra-red Paschen series in hydrogen, so valuable later to the Bohr theory, and in 1912 the discovery of the Paschen-Back effect in magneto-optics led to the removal of many anomalies and clarified the whole study of the Zeeman effect. The year 1916 saw, in a sense, the beginning of the study of hyperfine structure, for Paschen then investigated the structure of the ionized helium line 4686; and his data proved to be critically important in the early history of the Bohr theory of the atom, affording also a precision value for $\mathbf{e}/\mathbf{m}$. It was at this time, too, that Paschen invented the hollow cathode discharge, later to prove so fruitful for hyperfine structure studies in the hands of his pupils.

In 1919 appeared, what is now classic in spectroscopy, Paschen's classification of 130 series in the spectrum of neon. This was the first complex spectrum to be analysed and revealed a masterly grasp of a difficult subject. Three years later, Paschen-Götz's comprehensive report on spectral series and lists of classifications appeared. Simultaneously, Fowler's report on " Series in Line Spectra " was published by our Society, and both books were standards of reference for many years. Fowler and Paschen, although widely differing in temperament, were comparable as masters in elucidating complex spectra and their work overlapped. Both discovered isoelectronic sequences of stripped atoms, Fowler with silicon and Paschen with aluminium, and both announced their discovery in 1923. Paschen's spectroscopic notation was perhaps slightly better than that used by Fowler, and more closely resembles modern practice.

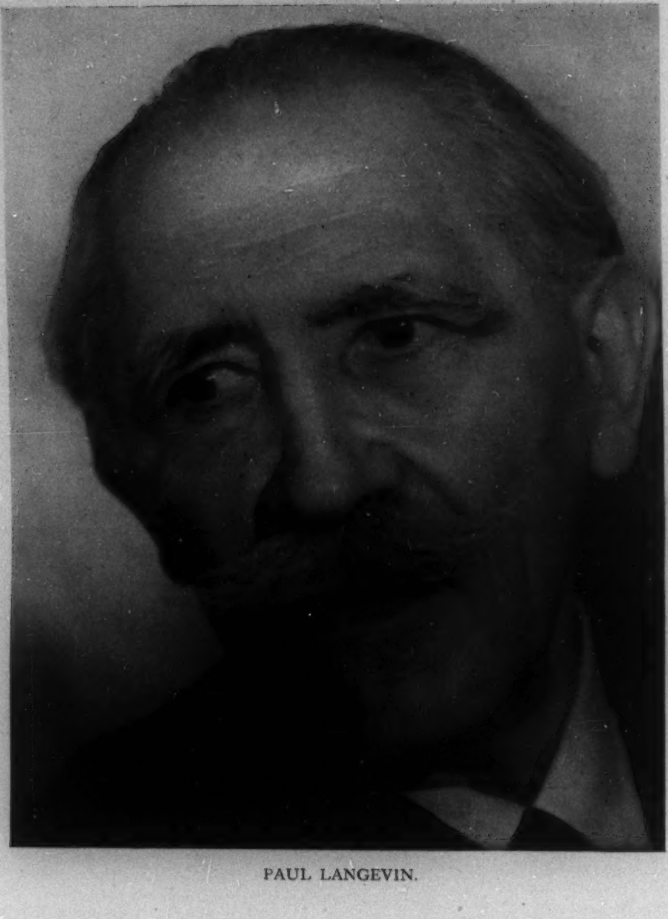
The next ten years brought a spate of publications, on series, Zeeman effect, thermal excitation, hyperfine structure, nuclear spin etc. despite very heavy administrative duties; for during this period he was the President of the Physikalisch-Technische Reichsanstalt, and this was no sinecure.

From 1931 onwards Paschen was violently attacked in the Nazi press, and he confided that this was due to his persistent purchase of scientific equipment abroad, for he always demanded the best of its kind. He was dismissed in 1933 and replaced by Stark. After 1933 he was permitted to use a small room in the Reichsanstalt and contented himsef, with measuring old plates and analysing them. From then on publication dwindleld but this was certainly not due to the weakening of old age, for the portrait shown here,



FRIEDERICH PASCHEN.

To face page 1040



PAUL LANGEVIN.

taken after 70, shows his virility. Withal he remained a strong, if silent, opponent of the Nazis, and in 1936 wrote a bitter letter in which he declared that spectroscopy in Germany was now destroyed, and lived only in England, U.S.A. and India.

In 1943 he lost practically all his possessions in a fire-raid, but despite his age kept on working. In the extreme cold of January 1947, accentuated by lack of fuel, he contracted pneumonia and died on 26 February in Potsdam.

Paschen was a man with immense vitality, and a much-loved photograph of him, appearing in the *Annalen der Physik* on his sixtieth birthday, shows a man simply exploding with energy. He made it a habit to devote at least an hour daily to discussion with research students, despite an overload of administration. That he was a great experimenter is clear from his record, but what was equally important was his infectious enthusiasm. All of his pupils, and they are many, were infected by his adventurous attack at a problem, and even at 70 he was still as excited as a boy, either by a new photograph, a pretty technical achievement, or a fundamental result.

He had a great respect for English spectroscopists, in particular, for A. Fowler and was a profound admirer of the late Lord Rayleigh, about whose writings in the *Encyclopaedia Britannica* he frequently talked. He was a Foreign Member of the Royal Society and an Honorary Fellow of the Physical Society.

He was a just man and very helpful to his pupils, in fact, an ideal director of research, inspiring, suggesting, criticizing and guiding, yet never taking any credit for his contributions to the discussion despite their real importance. He was lavish in praise of good work, and frankly condemned bad work with an expressive contemptuous sweep of the hand. He has trained a whole generation of spectroscopists who have much to thank him for.

S. TOLANSKY.

PAUL LANGEVIN

PROFESSOR LANGEVIN died peacefully in the early hours of Thursday morning 19 December 1946 at his residence in the École de Physique et de Chimie Industrielle, of which he was Director. He had been ill for some days, and in view of his age—he was 74—his death was not unexpected. His loss will be deeply felt in France, not only on account of his great services to science but also on account of his devotion to patriotic and democratic ideals.

Paul Langevin was born at Montmartre, Paris, on 23 January 1872, his parents being of Norman peasant stock. At the age of 16 he entered the École de Physique et de Chimie Industrielle, and after winning the diploma of this institute he became a student at the École Normale Supérieure. When he ended his studies there, in 1897, the city of Paris awarded him a scholarship to pursue his work at the Cavendish Laboratory at Cambridge. He remained there a year, under Sir J. J. Thomson. On his return to Paris he became demonstrator first at the École Normale and then at the Sorbonne, and joined in research work with Pierre and Marie Curie. He took his doctorate of science in 1902, in which year he became assistant professor at the Collège de France; in 1909 he was elected to the chair of general and experimental physics. He succeeded Pierre Curie in the School of Physics and Chemistry, of which he became director in 1925. In 1934 he became a member of the Académie des Sciences.

His research in physics, the scope of which was immense, included work on x rays, the properties of ions, the kinetic theory of gases, molecular orientation, and magnetic theory. It was while working at Cambridge under J. J. Thomson that he discovered the secondary rays of x rays. He was the first in France to make known Einstein's formula of relativity. His work on the mobility of ions and his identification of the large ions in the atmosphere is classic. So, too, is his investigation of magnetism on molecular theory, which was extended by Weiss. To the general public he was chiefly known for his discovery of supersonic waves and their application to the detection of submarines, work in applied physics inspired by the needs of war; but his influence on French physicists was greater than his actual work would suggest. It was in this connection that he studied and utilized the piezo-electric properties of quartz.

Langevin was keenly interested in the Solway Institute of Physics, a body whose efforts in stimulating the progress of theoretical physics between the two wars was of outstanding influence. On the death of H. A. Lorentz of Holland he succeeded to the Presidency of this Institute.

Langevin had long been interested in politics and was known for his advanced ideas. In 1939 he was the moving spirit of *La Pensée*, a serious Marxist quarterly, so that when he returned to Paris in October 1940, whence he had withdrawn with other officials of the Ministry of National Defence, to which he was temporarily attached, he was arrested by the Germans, who sent him to the Fresnes concentration camp. He was the first prominent figure of the French learned world to be incarcerated by the Germans. Protests from intellectuals in many countries led to his release; he was allowed to take up his work at Troyes, under close supervision. Meanwhile his son-in-law, Jacques Salomon, was arrested and shot by the Germans, and his daughter, Mme Salomon, was deported to Germany. In May 1944 his friends in the Resistance planned his escape to Switzerland. He came back to Paris in September of the same year, and at once joined the Front National and the Communist Party. In December he became president of the *Ligue des Droits de l'Homme*, and in April 1945 was elected municipal counsellor for Paris. Soon afterwards, the Ministry of National Education appointed him chairman of the commission for the reform of educational methods, and in June of this year scientific advisor to the commissariat for atomic energy. Recently he was chosen to be deputy chief French delegate to the UNESCO conference.

JOHN HENRY STRONG

In the early hours of 13 April 1947 disaster overtook the aircraft carrying the British observers to Araxa, Brazil, for the total eclipse of the Sun on May 20. Among those who sustained fatal injuries was J. H. Strong, Demonstrator in Astrophysics and Spectroscopy at the Imperial College of Science and Technology and a Student Member of the Physical Society. By this tragic accident at Dakar was cut short what promised to be a most fruitful scientific career.

Strong was born in London on 17 August 1924 and was educated at the Latymer Upper School, Hammersmith. Awarded a State Bursary in 1942, he came to Imperial College and graduated with first class honours in physics in 1944. He was then appointed a Demonstrator in the Physics Department and began research in spectroscopy. His first investigation, a study of the properties of a type of arc source designed for general spectroscopy analysis, was proceeding very satisfactorily and he had planned the publication of some of his results after returning from Brazil.

In performing an experiment, Strong was unusually observant and would undoubtedly have become an experimentalist of outstanding ability. His manner was quiet and gave little outward sign of the active and methodical way in which he was building up his knowledge of his chosen field of research. Both as an instructor and as a research colleague he was greatly liked and respected. His loss is deeply regretted.

R. W. B. PEARSE.

REVIEWS OF BOOKS

A Survey of the Principles and Practice of Waveguides, by L. G. H. Huxley. Pp. xi + 328. (Cambridge University Press, 1947). 21*s*.

During the war a great many fundamental advances were made in radio techniques as a result of the very rapid development of radiolocation (or radar as it now seems to be called) as a military weapon. These advances have been made known to the scientific public through the very successful Radiolocation Convention held in March last year under the auspices of the Institution of Electrical Engineers, and also through the published papers which have since appeared in the special issues (Part III A) of the Journal.* To many,

* Six special issues have now been published covering the lectures given at the Convention, and the supporting papers dealing with navigation and marine radar, aerials, waveguides, valves and cathode-ray tubes and receivers and transmitters.

however, these papers have been too specialized for easy study, and there has been a real need for a book, or books, providing an introductory survey of these developments. A series of books to fulfil this purpose has now been planned by the Cambridge University Press, and the present volume by Dr. Huxley dealing with the theory and practice of waveguides is the first of these to appear.

Dr. Huxley is Reader in Electromagnetism in the University of Birmingham, but, during the war, he was Head of the Radar School at the Telecommunications Research Establishment. He had thus a unique opportunity of becoming fully acquainted with all the recent developments, as well as learning at first hand the difficulties and pitfalls which present themselves in teaching the subject. We would therefore expect Dr. Huxley to write a clear and accurate account of the theory and practice of waveguides, in which the students' difficulties are appreciated and dealt with, and in this we are not disappointed. The treatment of the subject follows that of Dr. Huxley's courses at the Radar School. He deals with the properties and applications of waveguides, using simple physical explanations, and he relegates the formal mathematical treatment of field theory to the last chapter. The book appears to have lost nothing in content and vigour through this elementary and physical approach to the subject, and will certainly be appreciated by the less mathematical readers.

Chapter 1 gives an elementary treatment of the electromagnetic fields of T.E.M. waves. Chapter 2 discusses the general features of electromagnetic waves in metal waveguides and considers the various types of T.E. or H-waves and T.M. or E-waves and their relation to the T.E.M. wave. The chapter also includes a section dealing with the practical problem of launching these modes in circular and rectangular guides. Chapter 3 discusses the material of the earlier chapters in more mathematical language and evolves the formulae for the field components as functions of position in the guide. A discussion is also given of evanescent modes, piston attenuators, and attenuators of other types, as well as a consideration of attenuation due to the walls of the waveguides.

Chapters 4, 5 and 6 are decidedly the most interesting and probably the most valuable in the book. Chapter 4 discusses waveguide techniques, and deals with the special problems of the measurement of standing waves, the design of standing-wave indicators and their component parts (crystal detectors, crystal mixers etc.) reflectionless terminations, couplings, bends and corners, twists and tapers, and so on. All this material is well presented with full detail. § 4.12 dealing with E_0–H_0 transformers is a little unsatisfactory. The reviewer would have liked to have seen a much fuller account of these transformers.

Chapter 5 gives a general account of transmission line circuit theory using the concepts of characteristic impedance, line impedance, reflection coefficient etc., and then develops a parallel theory of waveguide transmission introducing the important concepts of waveguide impedance and normalized impedance. In all cases, however, where Dr. Huxley appeals to circuit theory, he is careful to point out the need for field concepts if a faithful description of the phenomenon is the aim. A physical interpretation of the normalized admittance or impedance of an obstacle or other discontinuity in a waveguide in terms of a scattering coefficient is given which should help in the understanding of these phenomena, and this is followed by a very full treatment of capacitive, inductive and resonant irises, etc. The treatment of T-junctions and their application to common aerial working T.R. systems, is excellent, but the section on waveguide switches is somewhat disappointing. The ring reflector switch and T.R. switch are quite rightly given special prominence, but little or nothing is said of other types, particularly variations of the ring switch, etc. The chapter also gives details of power measurements, the design of quarter-wave transformers and the design of corrugated waveguides. This latter is well presented, but very few details are given of the uses of corrugated waveguides and this section could well be amplified in the next edition. The chapter ends with a short section dealing with methods of feeding microwave aerials from waveguides. One slightly annoying feature of this and the previous chapter is the author's habit of occasionally referring by name to the originator of a particular idea or method. It would probably have been better to have omitted all reference to names of individuals (except, of course, when referring to published papers), or to have tried to mention all the chief contributors to the work. As it is, the author has omitted to mention the origin of many of the methods and techniques described, although the origin of many minor pieces of work is given in full.

Chapter 6 deals with cavity resonators and applications, and Chapter 7 attempts a formal treatment of field theory using single component Hertz vectors. This latter chapter contains much of interest and will well repay careful study.

The book is well printed and illustrated and appears to contain surprisingly few errors and mis-statements. It can be confidently recommended. D. TAYLOR.

Dissociation Energies and Spectra of Diatomic Molecules, by A. G. GAYDON, D.Sc. (Lond.), Warren Research Fellow of the Royal Society. Pp. xi + 239. (London: Chapman and Hall, Ltd., 1947.) 25*s*. net.

Dissociation energies of diatomic molecules, i.e. the energies required to dissociate such molecules from states without energy of translation, rotation or vibration, into states of two separate atoms at rest, are of the greatest theoretical interest, because they are the best measure of the chemical forces which bind two atoms together to form a molecule. They are also of great technical interest, e.g. in the investigation of combustion and of explosions.

The investigation of molecular spectra is a good example of the ever increasing co-operation of physicists and chemists. They can combine their efforts not only in the production of good spectra of pure materials, but also in the deciphering of the molecular spectra which are most complicated, and in their application to thermochemistry.

Dr Gaydon who has first edited an important book, *The Identification of Molecular Spectra*, with Dr. Pearse, and later an interesting book, *Spectroscopy and Combustion Theory*, publishes his new book from the standpoint of a thermochemist. He does not intend to introduce the reader to the theory of molecular spectra. This would be quite superfluous as we have an excellent introduction for physicists by G. Herzberg (*Molecular Spectra and Molecular Structure*) and for chemists by S. Glasstone (*Theoretical Chemistry*).

Gaydon gives only a very brief summary of the theoretical points which are relevant to the thermochemist (including such subjects as Potential Energy curves, Franck-Condon Principle and the correlations between molecular and atomic states), and concentrates mainly on his specific task to show first, by well investigated examples (O_2 and the halogen molecules), how the convergence limits of dissociation are found directly in molecular spectra and how the products of dissociation are determined. If the limits of convergence cannot be directly observed, this limit must be extrapolated. The method of extrapolation by Birge-Sponer is thoroughly discussed in Gaydon's book. The use which can be made of the phenomenon of predissociation for the determination of dissociation energies is treated in a separate chapter. Chapters are added on the thermal methods and the methods of controlled electron impacts which supplement the spectroscopic methods of determination of dissociation energies.

The most valuable part of the book consists in thorough *numerical* discussions of energies of dissociation of very important substances, especially of N_2, N_2^+, NO, CO, CO^+, CN, C_2N_2, and in a collection of the numerical data of about 250 diatomic molecules. Such critical data with all references from scientific periodicals and standard works can only be given by an expert and represent scientific work in its most condensed form.

The book which includes 39 figures and 4 plates of spectrograms, is written in a clear and stimulating style. It will be of great value for physicists as well as chemists (especially for thermochemists) and can be warmly recommended.

K. JELLINEK.

Fourier Transforms and Structure Factors, by DOROTHY WRINCH. Pp. ix + 96. (Asxred Monograph No. 2, 1946.) $4.00.

In a review of a scientific work it is often assumed that the reader is familiar with the subject matter, the reviewer restricting himself to the presentation of his views as to how well the author has succeeded in achieving his object, with hints, in some cases, on how it might have been done better. Occasionally, however, a book comes along on a subject with which the reader will almost certainly not be familiar, and in this case the reviewer's function becomes more that of an interpreter. Such a book is the monograph *Fourier Transforms and Structure Factors*, by Dorothy Wrinch, which arises from a long-term research on the structure of the native proteins, and which will certainly not be wholly understandable by every X-ray crystallographer, but which may well prove to be of great importance in structure analysis.

Briefly, Dr. Wrinch sets out to show how the structure factor of a particular atomic grouping may be combined with those of other atomic groupings to determine the structure factor of a crystal as a whole, so that a knowledge of the structure factors of frequently occurring atomic groupings may be used whenever the appropriate groupings arise, or are thought to be capable of arising, without recourse to the consideration of individual atomic positions. Difficulties on account of the variation of scattering power and atomic size are overcome by working throughout with the " transform " of a distribution (the structure factor per effective electron) rather than with the structure factor itself, and by plotting the values of these transforms in the reciprocal lattice on an arbitrary scale, the units of which can be varied appropriately to suit any particular case.

The transforms of a variety of frequently occurring groups—the vertices of a cube, of a tetrahedron and of an octahedron, and the "benzene" point set—are worked out, and it is shown how the transform concept can be extended to include Patterson distributions. From these relatively simple transforms Dr. Wrinch goes on to consider transforms of many points arranged in patterns on surfaces or in depth, and finally considers the transforms of various distributions in continuous space, rather than periodic distributions such as occur in crystals. A chapter is added on the application of the transform concept to the study of small crystals, and there is an appendix on Fourier series and Fourier's integral theorem.

In the preface, Dr. Wrinch makes the point that a systematic study of the " language of structure factors " is a necessary preliminary to the interpretation of the intensity maps of crystals made up of megamolecules of unknown structures. The present monograph will render such a study much easier and should certainly do something to encourage the use of transforms in ordinary crystal analysis, a use which has so far been very restricted.

J. THEWLIS

Sound: *A Physical Textbook*, by E. G. RICHARDSON, D.Sc. 4th edition. Pp. 344. (Edward Arnold and Co., 1947.) 18*s*.

Since the appearance of the first edition of Dr. Richardson's book in 1927 much water has flowed under London Bridge, or perhaps one might say more appropriately " much sound has degenerated into heat ". Much of this sound served a useful purpose but we know full well that much of it did not!

In the preface to the first (1927) edition Dr .Richardson referred to the renewed attention to this branch of physics " stimulated no doubt by the European War (1914–18) and by the development of broadcasting ". If further stimulus were required this has no doubt been provided by the second war. Dr. Richardson's well known book has served a very useful purpose between the two wars in encouraging the study of that very interesting subject of " Sound ". His book has now reached the fourth edition after a period of 20 years since its first appearance. In the latest edition the whole subject matter has been re-arranged and brought up to date. It is fitting to quote from the preface of this edition: " The position in sound is now very different from what it was when this book was first written: there are a number of books dealing with special aspects of applied acoustics and at the outbreak of the second world war there were three periodicals and one society dealing solely with the subject, whilst in industry the penetration of applied acoustics has been far-reaching ". It may be of interest to add also that the Physical Society has recently formed an Acoustics Group which has already a very considerable membership both of " academic " and ' industrial ' scientists and engineers.

Dr. Richardson covers the general field of sound and, the reviewer is pleased to note, is now taking a little more interest in under-water sound. The chapters on " Vortex Formation and Jet Tones " and on " Columns of Air " are particularly noteworthy, as no other textbook of sound known to the reviewer deals so thoroughly with these important branches of the subject. They are all the more valuable as Dr. Richardson himself has played a leading part in the work which he describes. Referring again to under-water sound, there is a little confusion on p. 324 in the description of the construction and mode of operation of two hydrophones, one of the 'light-body', 'displacement type' (illustrated) and the other of the metal diaphragm 'pressure type'.

The book is very readable and covers a wide field in a relatively small compass. The appearance of the fourth edition is sufficient testimony to its continued popularity and usefulness. This excellent textbook needs no further recommendation. A. B. WOOD.

German-English Science Dictionary, by DE VRIES. Second edition, revised and enlarged; second impression. Pp. xiv+558. (London: McGraw-Hill Publishing Co., 1947.) 22*s.* 6*d.*

This is a dictionary intended for students in the physical, biological and related sciences. Originally containing 48,000 entries, this new edition contains additional words and idioms and incorporates a great number of suggestions made by users of the first.

The gender and plural form of each noun are given explicitly, except for nouns belonging to the large class of regular feminine plurals. One most welcome feature in a dictionary of this kind, which in the main has users who are not language specialists, is the printing of the parts of strong verbs as separate entries. It is thus not necessary, for example, to recognize *schlägt* or *schlug* as parts of the verb *schlagen* before one can find their meanings. The infinitive of the verb is given in brackets after each part so printed. This should prove of considerable assistance to students (and others) for whom languages are not a very strong point.

Compound nouns are, in general, not entered as such, the reader being expected to look up the different parts of such words separately. It must be said, however, that this rule is very often not adhered to, even in cases where the compounding has introduced no special meaning. Thus one finds *Messergebnis*, *Messfehler*, etc., entered separately—and such examples occur on almost every page. However, to most users this will be looked upon rather as an advantage.

A weakness, and one which tends to mar this otherwise excellent dictionary, is the failure to distinguish compound verbs in their separable and inseparable forms. This is serious, since the same verb used separably may differ completely in meaning from the inseparable form. Thus *über/legen*=to lay over; but used inseparably *überlegen*=to consider. This verb is, in fact, entered only once—and these two widely differing meanings are apparently given as synonyms! It is to be hoped that this will be corrected in future editions.

This last criticism should not be allowed to detract from the fact that in other respects this dictionary offers real assistance to anyone who has occasion to read scientific German—and that includes a large enough body of workers to say that many people will find this work helpful and some invaluable. H. H. HOPKINS.

Guide to the Literature of Mathematics and Physics, by NATHAN GREER PARKE. III. Pp. xv+205. (New York and London: McGraw-Hill Book Co. Inc., 1947.) $5.00.

This book falls into two parts. About two-fifths is occupied with a general guide in which the respective uses of handbooks, text-books, encyclopedias and dictionaries are set out, methods of study discussed, methods of literature search described, indexing and cataloguing shortly treated, the main abstracting journals mentioned, and so on. This material is quite interesting and helpful, but differs entirely from that in the second part of the book. In the first part, just described, the material is for study and assimilation, whereas the second part is simply, and valuably, for reference.

The second part contains, in classified sections, the titles and bibliographical particulars of some 2300 books likely to be of use at one time or another to any physicist. The three main fields, of mathematics, physics and engineering, are each split up into a few broad areas (9 for physics) and each of these into perhaps a dozen sub-areas, of which there are 150 altogether. These 150 sub-areas are then given in alphabetical order (so that sub-areas from different broad ones come together) and under each alphabetical entry is given a short commentary and a list of text-books or sources which might be consulted on it.

These books are not confined to those from any one country; French and German, American and British, all appear. It would not be appropriate to discuss in detail the selection actually made—it is a personal one, and the reviewer's, whilst it might differ, would be equally personal. It may be of interest, however, to prospective purchasers, to know what sort of thing he is likely to find, and we may fairly note, therefore, that in some cases early sources are included, but that this is not always so. Thus, all the five books on algebraic equations are published in America and dated 1938 or later, so that Cayley, Sylvester and Salmon do not appear, and the treatise by Burnside and Panton is not listed. On the other hand, under " Elliptic functions ", we find the *fundamenta nova* of Jacobi (1829) but not Lagrange's *Traité*. Here, again, we find a standard English treatise, that of Cayley (1895) but not the very useful little book by Dixon.

It may, then, be seen that here is a useful reference book which may be very helpful when one is seeking the literature of a subject new to one, and may be a pleasant browsing companion at other times. It is as well produced as the McGraw books generally are.

J. H. A.

The Principles of Quantum Mechanics, by P. A. M. DIRAC. 3rd edition. Royal 8vo. Pp. 311. (Oxford: The University Press, 1947.) 25*s*. net.

It is not often that a book written within five years of the first introduction of an entirely new theory, so wide and so fundamental as quantum theory and wave mechanics, remains the standard textbook for more than fifteen years. Yet this is the case with Dirac's *Principles of Quantum Mechanics*. It was as long ago as 1930, less than five years after the introduction of Heisenberg's matrix mechanics, and Schrodinger's wave mechanics, that the first edition of this book appeared, and the late Sir Arthur Eddington wrote of it that "for the first time wave mechanics is presented in a really coherent form with something like a philosophy of the new methods to support it ". The fact that so few fundamental changes have been found necessary in later editions (of which this is the third) is proof enough that Eddington was right. Though the book is not everywhere easy to read, no one who is genuinely interested in the general formulation of quantum theory can afford to neglect it.

Substantially this third edition closely resembles the second edition. The theory of quantum electrodynamics has advanced a little (but only a little) since 1935, so that there is now found room for the Wentzel field, and the λ-limiting process. As Dirac says, to go further would be to trespass " on speculative ground ". A more powerful method of dealing with systems of like particles is used, based on Fock's treatment of the theory of radiation. But the most noticeable change, at least to the eye, is the introduction of the *bra* and *ket* notation. What was previously called ψ, the representative of a state, is now called a *ket*, and written $|\,\rangle$; the dual vector, hitherto called ϕ, is now called a *bra*, and written $\langle\,|$. The product of the two vectors $\langle B|$ and $|A\rangle$, which is always a pure number, and which appears very frequently, is no longer $\phi\psi$, but $\langle B|A\rangle$. The use of *bra* and *ket* is obvious when we decide to call this latter expression a " bracket ", or " complete bracket ", compared with which the others are " incomplete brackets ". The change of notation has the advantage of greatly simplifying some of the formalism.

As usual the Oxford University Press has turned out a first-rate piece of printing.

C. A. C.

Applied Bessel Functions, by F. E. RELTON. Pp. vi+191, 10 figures. (Blackie and Son, 1947.) 17*s*. 6*d*.

This book, written by an applied mathematician, is very unconventional in the order of presentation it adopts. The author is at pains to justify this approach in his preface—and draws on the support of E. B. Wilson in this connection. He had in fact little need of this, for his text provides more than adequate justification.

The material of the book is presented in a uniformly pleasant (almost chatty) style, which would perhaps be felt obtrusive in a work of bigger proportions. Here, however, in a book intended for the new physics or engineering graduate, it seems quite in place. What is more it enables the author to point a few morals to the tyro in applied mathematics which a more formal style might well have inhibited. As a consequence these interpolations make the book both more interesting and more informative.

The first chapter is devoted to a preliminary study of the error function, and of the gamma and beta functions, while Chapter 2 reviews those parts of the theory of linear differential equations which are necessary to later parts of the book. In Chapter 3 a class of functions is defined by means of recurrence formulae. It is shown that these functions are solutions of Bessel's equations, and are accordingly Bessel functions. The series solution of Bessel's equation is not given until Chapter 4; and Chapter 5 then deals with applications involving the Bessel functions of the first kind, to which attention is at first confined.

Chapters 6 and 7 see the introduction of the Bessel functions of the second kind, the modified Bessel functions, and further applications calling for their use. The applications to hydrodynamics and elasticity, to which Chapter 8 is devoted, will probably be found the most difficult for the beginner. The various integrals and expansions involving Bessel functions are not treated until Chapter 9—in itself sufficient to evidence the unconventional presentation of the subject. The final chapter deals very briefly with allied functions and asymptotic expansions.

For anyone anxious to know something about Bessel functions there are books as good as this, but few could claim to be better. H. H. HOPKINS.

Photographic Recording of Cathode-ray Tube Traces, by R. J. HERCOCK. Pp. 60. (London: Ilford Ltd., 1947.) 5*s*.

The book forms " No 1 of a series of Ilford technical monographs ", and " It is the object of this booklet to provide those familiar with the use of cathode-ray tubes with an insight into the photographic technique necessary to obtain useful records of traces ". The book does not, however, deal solely with the photographic aspects. It also describes tubes and methods of application. The author contrives to cover a wide field in the short space of sixty pages by adopting a pleasantly terse style and by approaching the reader at a fairly high technical level. Thus, certain photographic terms, such as " density " are used but not explained, as is indeed unnecessary. The book thus assumes some familiarity both with physical and photographic terms.

Photography as a scientific tool is treated as a Cinderella in some laboratories. For this reason, the book will form a welcome addition in the libraries of many institutions and will fill a gap among existing text-books.

The scope of the book does not extend beyond published material. Insufficient stress is laid on the necessity for full development with high recording speeds and thus short exposures; there is also no mention of recent work on latent image intensification, which is often useful in this field. There is one peculiar misprint: a density of 0·1 is quoted as corresponding to a change of 12½ % in transmission, the correct figure being 26 %. The book suffers but slightly from the naturally exclusive preoccupation with Ilford materials, and the scientific world will look forward to further volumes in this series.

W. F. B.

Light, Vision and Seeing, by MATTHEW LUCKIESH. Pp. xiv + 323. New York: D. Van Nostrand Company, Inc., 1944.) 25*s*. net.

In an earlier work—*The Science of Seeing*—Dr. Luckiesh and Mr. Moss recapitulated the valuable series of investigations on visual capacity, lighting and human welfare which had been carried out over some thirty years by the authors and their colleagues at the Research Laboratories of the General Electric Company, Cleveland. The present book, addressed to readers interested but not necessarily knowledgeable in the subject, gives a simplified and detailed exposition of the more elementary topics of the earlier work, states the principal conclusions and includes additional matter on the historical development of lighting methods and their impact on everyday life. While the English reader may find the style a little florid in places, there is no doubt that Dr. Luckiesh has produced an easily understood and very informative account of the relation between light and sight. There are a few slips of the pen, for example the statement on p. 287 that the molecules of air are comparable in size with the wave-lengths of light. W. S. S.

INDEX

PAGE

Absorption and selective photo-effect in adsorbed layers 30
Acceleration of charged particles to very high energies 666
Acceleration of electrons, Multiple-gap linear 622
Adiabatic temperature changes and magnetization of cobalt 329
Aircraft camera lenses; discussion (**58**, 493, 1946) 155
Allen, W. D. and Symonds, J. L.: Experiments in multiple-gap linear acceleration of electrons 622
Ammonia: Collision broadening of the inversion spectrum at centimetre wavelengths: I 418
Andrews, J. P.: Thermoelectric power of cadmium oxide; and discussion . . 990
Aplanatic lenses for unit magnification 844
Appleton, Sir Edward, and Beynon, W. J. G.: The application of ionospheric data to radio communication problems: Part II; and discussion . . . 58, 534
Appleton, Sir Edward, and Naismith, R.: The radio detection of meteor trails and allied phenomena 461
Arc, copper, The effect at the cathode 273
Aspheric profiles, On the determination of 704
Auditorium, Mean free path of sound in 535
Avery, D. G. and Witty, R.: Diffusion pumps: A critical discussion of existing theories 1016

β-spectroscopes, prism, The design of 905
Band spectra of CS and CSe 107
Band-spectroscopic data, The calculation of potential-energy curves from . . 998
Band-systems, ultra-violet absorption, of PbO, PbS, PbSe and PbTe . . . 449
Bands of Na_2, Ultra-violet 610
Barrow, R. F., *see* Vago, E. E.
Bate, A. E. and Pillow, M. E.: Mean free path of sound in an auditorium . . 535
Bates, L. F. and Edmondson, A. S.: The adiabatic temperature changes accompanying the magnetization of cobalt in low and moderate fields 329
Bates, W. J.: A wavefront-shearing interferometer 940
Beynon, W. J. G.: Oblique radio transmission in the ionosphere, and the Lorentz polarization term; and discussion 97, 534
Beynon, W. J. G.: Observations of the maximum frequency of radio communication over distances of 1000 km. and 2500 km.; and discussion . . . 521, 534
Beynon, W. J. G., *see* Appleton, Sir Edward.
Bleaney, B., Loubser, J. H. N. and Penrose, R. P.: Cavity resonators for measurements with centimetre electromagnetic waves 185
Bleaney, B. and Penrose, R. P.: Collision broadening of the inversion spectrum of ammonia at centimetre wavelengths: I 418
Bragg x-ray spectrometer, Optical model demonstrating 111
Brossel, J.: Multiple-beam localized fringes 224, 234
Bromine, Interpretation of the visible spectrum of 1008
Brown, R. C.: The fundamental concepts concerning surface tension and capillarity; and discussion, and corrigenda 429, 711
Brunt, Professor D., President of the Physical Society, 1945–1947; Portrait *frontispiece*
Brunt, D.: Some physical aspects of the heat balance of the human body. (Presidential Address) 713
Bryan, G. B. (Obituary notice) 508
Burch, C. R.: Reflecting microscopes 41, 47
Butler, C. C., *see* Rymer, T. B.

PAGE
Cadmium oxide, Thermoelectric power of 990
Camera lenses, Aircraft; discussion (**58**, 493, 1946) 155
Capillarity and surface tension, Fundamental concepts concerning; and corrigenda 429, 711
Cathode of the copper arc, A note on the effect at 273
Cavity resonators for measurements with centimetre electromagnetic waves 185
Centimetre electromagnetic waves, Cavity resonators for measurements with 185
Centimetre wavelengths, Collision broadening of the inversion spectrum of ammonia at: I 418
Centimetric electromagnetic waves over ground, Reflection and diffraction effects with wire-netting screens 847
Charged particles to very high energies, Acceleration of 666
CO, Photo-chemical decomposition of; addendum to discussion 502
Coaxial electron lenses 828
Cobalt, The magnetization and adiabatic temperature changes of 329
Collision broadening of the inversion spectrum of ammonia at centimetre wavelengths: I 418
Colorimeter with six matching stimuli 554
Colorimetry in the glass industry 592
Coloured light signals near the limit of visibility, Recognition of 560
Coloured point sources against a white background, Measurement of the chromatic and achromatic thresholds of 574
Colours, Distribution coefficients for the calculation of, on the C.I.E. trichromatic system for total radiators 814
Corrigenda 33, 711, 901
Coutts, W. B. (Obituary notice) 508
Cowley, J. M. and Rees, A. L. G.: Refraction effects in electron diffraction 287
Craggs, J. D. and Hopwood, W.: Electron/ion recombination in hydrogen spark discharges 771
Craggs, J. D. and Hopwood, W.: Ion concentrations in spark channels in hydrogen 755
Craig, H.: The production of a uniform magnetic field over a specific volume by means of twin conducting circular coils 804
Crystal structure of gold leaf, Determination by electron diffraction 541
CS and CSe, The spectra of 107
Cubic lattice, Dislocations in a simple 256
Cuer, P., *see* Lattes, C. M. G.

Daly, E. F. and Sutherland, G. B. B. M.: An infra-red spectroscope with cathode-ray presentation; and discussion 77, 901
Damping capacity, strain hardening and fatigue 275
Damping, radiation, theory of 917
Diffraction effects with wire-netting screens, and reflection of centimetric electromagnetic waves over ground 847
Diffusion, The hole theory of 694
Diffusion pumps: a critical discussion of existing theories 1016
Dislocations in a simple cubic lattice 256
Domb, C., The theory of an oscillator coupled to a long feeder, with applications to experimental results for the magnetron 958
Donaldson, R.: A colorimeter with six matching stimuli 554
Dungey, J. W. and Hull, Catherine R.: Coaxial electron lenses 828
Dymond, E. G.: The Kew radio sonde 645

Edmondson, A. S., *see* Bates, L. F.
Edmondson, A. S.: The matching of high frequency transmission lines using a frequency-variation method 982
Eisenschitz, R.: The effect of temperature on the thermal conductivity and viscosity of liquids 1030
Electro-acoustic transducers, Sensitivity and impedance of 19
Electromagnetic centimetre waves, Cavity resonators for measurements with 185
Electron diffraction, Refraction effects in 287

PAGE

Electron/ion recombination in hydrogen spark discharges 771
Electron lenses, Coaxial 828
Electron optics and space charge in strip-cathode emission systems . . . 302
Electrons, Experiments in multiple-gap linear acceleration of 622
Elliott, H. A.: An analysis of the conditions for rupture due to Griffith cracks . 208
Emissivity of hot metals in the infra-red 118, 131
Evans, J. C.: The determination of thermal lagging times; and discussion . 242, 1039

Fatigue, damping capacity, strain hardening and 275
Faust, R. C. and Tolansky, S.: A transparent-replica technique for interferometry 951
Flicker noise in valves and impurity semi-conductors; and discussion . . 366, 403
Fluctuations in streams of thermal radiation 34
Fluctuations, spontaneous, of electricity in thermionic valves; and discussion 375, 388, 403
Fowler, P. H., *see* Lattes, C. M. G.
Fracture in glass, Delayed; and discussion 169, 1036
Frank, F. C.: The mass of the neutrino 408
Freezing-in of nuclear equilibrium 139
Fröhlich, H. and Sack, R. A.: Light absorption and selective photo-effect in adsorbed layers 30
Fundamental physics, What experiments are needed in (lecture and discussion) . 412
Fürth, R. and MacDonald, D. K. C.: Spontaneous fluctuations of electricity in thermionic valves; and discussion 375, 388, 403

Generator, High-voltage, at Imperial College 699
Gerö, L., *see* Schmid, R. F.
Glass, Delayed fracture in; and discussion 169, 1036
Glass industry, Colorimetry in the 592
Glückauf, E.: Investigations on absorption hygrometers at low temperatures . 344
Gogate, D. V. and Pathak, P. D.: The Landau velocity in liquid helium: II . 457
Gold leaf, Determination of the crystal structure by electron diffraction . . 541
Gooden, J. S., *see* Oliphant, M. L.
Gooden, J. S., Jenson, H. H. and Symonds, J. L.: Theory of the proton synchroton; and corrigenda 677, 901
Griffith's theory of rupture, Extension of, to three dimensions; discussion (**58**, 729, 1946) 1036
Griffith cracks, An analysis of the conditions for rupture due to 208
Grimmett, L. G., *see* Mann, W. B.
Gurney C.: Delayed fracture in glass; and discussion 169, 1036
Gurney, C.: Thermodynamic relations for two phases containing two components in equilibrium under generalized stress 629

Hamilton, J.: The theory of radiation damping 917
Hanstock, R. F.: Damping capacity, strain hardening and fatigue 275
Harding, H. G. W. and Sisson, R. B.: Distribution coefficients for the calculation of colours on the C.I.E. trichromatic system for total radiators at 1500–250–3500° K., and 2360° K. ($C=14\,350$) 814
Heat balance of the human body, Some physical aspects of the 713
Heat transfer, Some investigations in the field of 726
Heavy elements in stars, The formation of 972
Helium II, liquid, Landau velocity in 457
Hey, J. S., Parsons, S. J. and Jackson, F.: Reflection of centimetric electromagnetic waves over ground, and diffraction effects with wire-netting screens . . 847
Hey, J. S. and Stewart, G. S.: Radar observations of meteors 858
Hide, G. S., *see* Oliphant, M. L.
High-frequency transmission lines, matching by a frequency-variation method . 982
Hill, N. E. G.: The measurement of the chromatic and achromatic thresholds of coloured point sources against a white background 574
Hill, N. E. G.: The recognition of coloured light signals near the limit of visibility 560

PAGE
Hole theory of diffusion 694
Holmes, J. G.: Colorimetry in the glass industry 592
Hopwood, W., *see* Craggs, J. D.
Howell, H. G.: The spectra of CS and CSe 107
Hoyle, F.: The formation of heavy elements in stars 972
Hull, Catherine R., *see* Dungey, J. W.
Human body, Some physical aspects of the heat balance of 713
Hydrogen, Ion concentrations in spark channels in 755
Hydrogen spark discharges, Electron/ion recombination in 771
Hydrostatic tension, The behaviour of water under, III 199
Hygrometers, absorption, at low temperatures 344

Impedance and sensitivity of electro-acoustic transducers 19
Imperial College high-voltage generator 699
Inclined plane pole-faces, Magnetic focusing between 791
Infra-red spectroscope with cathode-ray presentation; and discussion . . 77, 901
Infra-red, The emissivity of hot metals in the 118, 131
Interferometer, A wavefront-shearing 940
Interferometry, A transparent-replica technique for interferometry 951
Intermittency effect, A note on the 161
Ion concentrations in spark channels in hydrogen 755
Ionosphere, Oblique radio transmission in, and the Lorentz polarization term; and discussion 97, 534
Ionospheric data, The application to radio communication problems: II; and discussion 58, 534
Ionospheric region, Equivalent path and absorption in an 87

Jackson, F., *see* Hey, J. S.
Jaeger, J. C.: Equivalent path and absorption in an ionospheric region . . . 87
Jakob, Max: Some investigations in the field of heat transfer 726
Jeans, Sir James (Obituary notice) 503
Jensen, H. H., *see* Gooden, J. S.
Jets, liquid, The break-up of 1

Kew radio sonde 645
Klemperer, O.: Electron optics and space charge in strip-cathode emission systems 302

Laby, T. H. (Obituary notice) 506
Lagging times, The determination of thermal 242
Landau velocity in liquid helium II 457
Langevin, P. (Obituary notice) 1041
Latham, R.: Nuclear magnetic moments 979
Lattes, C. M. G., Fowler, P. H. and Cuer, P.: A study of the nuclear transmutations of light elements by the photographic method 883
Lenses, Aircraft camera; discussion 155
Lenses, aplanatic, for unit magnification 844
Lewis, W. B.: Fluctuations in streams of thermal radiation 34
Lines of force through neutral points in a magnetic field 14
Liquid helium II, Landau velocity in 457
Liquid jets, The break-up of 1
Liquids, The effect of temperature in the thermal conductivity and viscosity of . 1030
Lord, Mary P., Rees, A. L. G. and Wise, M. E.: The short-period time variation of the luminescence of a zinc sulphide phosphor under ultra-violet excitation; and corrigenda 473, 711
Loubser, J. H. N., *see* Bleaney, B.
Luminescence of a zinc sulphide phosphor under ultra-violet excitation, Short-period time variation of; and corrigenda 473, 711

MacDonald, D. K. C., *see* Fürth, R.
Macfarlane, G. G.: A theory of flicker noise in valves and impurity semi-conductors; and discussion 366, 403

PAGE
Magnetic field, Lines of force through neutral points in 14
Magnetic field, Uniform production over a specific volume by means of twin conducting circular coils 804
Magnetic focusing between inclined plane pole-faces 791
Magnetic moments, nuclear 979
Magnetic prisms, The optical properties of axially symmetric 905
Magnetization of cobalt in low and moderate fields, The adiabatic temperature changes accompanying 329
Magnetron, Application of theory of an oscillator and long feeder to experimental results for 958
Mann, W. B. and Grimmett, L. G.: The Imperial College high-voltage generator. 699
Maximum frequency of radio communication over distances of 1000 km. and 2500 km.; and discussion. 521, 534
May, J.: corrigendum 33
Merrington, A. C. and Richardson, E. G.: The break-up of liquid jets . . 1
Meteor trails and allied phenomena, The radio detection of 461
Meteors, Radar observations of 858
Micrometer, time, of high accuracy 585
Microscopes, Reflecting 41, 47
Milbourn, M.: The effect at the cathode of a copper arc 273
Multiple-beam localized fringes 224, 234
Multiple-gap linear acceleration of electrons, Experiments in 622

Na_2, Ultra-violet bands of 610
Nabarro, F. R. N.: Dislocations in a simple cubic lattice 256
Naismith, R., *see* Appleton, Sir Edward.
Neutrino, The mass of 408
Neumann, E. A.: A time micrometer of high accuracy 585
Nickel, The variation of the reflectivity with temperature 781
Nuclear equilibrium, The freezing-in of 139
Nuclear magnetic moments 979
Nuclear transmutations of light elements, A study by the photographic method . 883

Obituary notices 503, 1040
Oliphant, M. L.: Rutherford and the modern world. (Third Rutherford lecture). 144
Oliphant, M. L., Gooden, J. S. and Hide, G. S.: The acceleration of charged particles to very high energies 666
Optical properties of axially symmetric magnetic prisms 905
Oscillator coupled to a long feeder, with application to results for the magnetron . 958
Owen, D.: The lines of force through neutral points in a magnetic field; and corrigendum 14, 901

Parsons, S. J., *see* Hey, J. S.
Paschen, F. (Obituary notice) 1040
Paterson, S.: The heating or cooling of a solid sphere in a well-stirred fluid . 50
Pathak, P. D., *see* Gogate, D. V.
PbO, PbS, PbSe and PbTe, Ultra-violet absorption band-systems of 449
Peierls, R. E.: What experiments are needed in fundamental physics? (lecture and discussion) 412
Penrose, R. P., *see* Bleaney, B.
Permittivity, effective, of two-phase systems 1011
Phosphor, zinc sulphide, Short-period time variation of the luminescence of; and corrigenda 473, 711
Photo-chemical decomposition of CO; addendum to discussion 502
Photo-effect and light absorption in adsorbed layers 30
Pillow, M. E., *see* Bate, A. E.
Preddy, W. S., *see* Wolf, E.
Price D. J.: The emissivity of hot metals in the infra-red 118, 131
Prisms, magnetic, axially symmetric, Optical properties of 905

PAGE

Profiles, aspheric, On the determination of 704
Proton synchroton, Theory of the 677
Pumps, diffusion : a critical discussion of existing theories 1016

Radar observations of meteors. 858
Radiation damping, The theory of 917
Radiators, total, Distribution coefficients for the calculation of colours on the C.I.E. trichromatic system for 814
Radio communication, the maximum frequency of, over distances of 1000 km. and 2500 km.; and discussion 521, 534
Radio communication problems : The application of ionospheric data to : II ; and discussion 58, 534
Radio detection of meteor trails and allied phenomena 461
Radio sonde, The Kew 645
Radio transmission, oblique, in the ionosphere, and the Lorentz polarization term; and discussion 97, 534
Rees, A. L. G., *see* Cowley, J. M.
Rees, A. L. G., *see* Lord, Mary P.
Rees, A. L. G. : The calculation of potential-energy curves from band-spectroscopic data 998
Rees, A. L. G. : Note on the interpretation of the visible absorption spectrum of bromine 1008
Reflecting microscopes 41, 47
Reflection of centimetric electromagnetic waves over ground, and diffraction effects with wire-netting screens 847
Reflectivity of nickel, The variation of, with temperature 781
Refraction effects in electron diffraction 287
Replica, transparent, technique for interferometry 951
Resonators, cavity, for measurements with centimetre electromagnetic waves . 185
Reviews of books 157, 326, 509, 711, 902, 1042
Richardson, E. G., *see* Merrington, A. C.
Richardson, H. O. W. : Magnetic focusing between inclined plane pole-faces . 791
Rupture, Extension of Griffith's theory of, to three dimensions; discussion (**58**, 729, 1946) 1036
Rupture due to Griffith cracks, An analysis of the conditions for 208
Rushman, D. F. and Strivens, M. A. : The effective permittivity of two-phase systems 1011
Rutherford and the modern world (Third Rutherford lecture) 144
Rymer, T. B. and Butler, C. C. : Determination of the crystal structure of gold leaf by electron diffraction 541

Sack, R. A. : Extension of Griffith's theory of rupture to three dimensions; discussion (**58**, 729, 1946) 1036
Sack, R. A., *see* Fröhlich, H.
Schmid, R. F. and Gerö, L. : Addendum to discussion 502
Selwyn, E. W. H. and Tearle, J. L. : The performance of aircraft camera lenses; discussion (**58**, 493, 1946) 155
Semi-aplanat reflecting microscopes 47
Semi-conductors, impurity, A theory of flicker noise in valves and; and discussion 366, 403
Sensitivity and impedance of electro-acoustic transducers 19
Siday, R. E. : The optical properties of axially symmetric magnetic prisms : I ; and discussion 905
Sinha, S. P. : Ultra-violet bands of Na_2 610
Sisson, R. B., *see* Harding, H. G. W.
Smith, T. : A series for the stationary value of a function 323
Smith, T. : Note on aplanatic lenses for unit magnification 844
Sound in an auditorium, Mean free path of 535
Space and charge in strip-cathode emission systems, Electron optics and . . 302
Spark channels in hydrogen, Ion concentrations in 755

PAGE
Spark discharges, hydrogen, Electron/ion recombination in 771
Spectra of CS and CSe 107
Spectra, ultra-violet absorption, of PbO, PbS, PbSe and PbTe 449
Spectroscope, infra-red, with cathode-ray presentation; and discussion. . 77, 901
Spectrum of bromine, visible absorption, Interpretation of 1008
Spectrum of Na_2 610
Sphere, solid, The heating or cooling in a well-stirred fluid. 50
Stars, The formation of heavy elements in 972
Stationary value of a function, A series for the 323
Stewart, G. S., *see* Hey, J. S.
Strain hardening, Damping capacity and fatigue 275
Stress, generalized, Thermodynamic relations for two phases containing two components in equilibrium under 629
Strivens, M. A., *see* Rushman, D. F.
Strong, J. H. (Obituary notice) 1042
Surface tension and capillarity, Fundamental concepts concerning; and discussion 429, 711
Sutherland, G. B. B. M., *see* Daly, E. F.
Symonds, J. L., *see* Allen, W. D.
Symonds, J. L., *see* Gooden, J. S.
Synchrotron, proton, Theory of the 677

Tearle, J. L., *see* Selwyn, E. W. H.
Temperley, H. N. V.: The behaviour of water under hydrostatic tension: III . . 199
Thermal conductivity and viscosity of liquids, effect of temperature on. . . 1030
Thermal lagging times, The determination of 242
Thermal radiation, Fluctuations in streams of 34
Thermodynamic relations for two phases containing two components in equilibrium under generalized stress 629
Thermoelectric power of cadmium oxide 990
Thermionic valves, Spontaneous fluctuations of electricity in; and discussion 375, 388, 403
Time micrometer of high accuracy 585
Tolansky, S., *see* Faust, R. C.
Transducers, electro-acoustic, Sensitivity and impedance of 19
Transmutations, nuclear, of light elements, A study by the photographic method . 883
Transparent-replica technique for interferometry 951

Ubbelohde, A. R.: The freezing-in of nuclear equilibrium 139
Uranium chain reaction, A mechanical model illustrating the 113

Vago, E. E. and Barrow, R. F.: Ultra-violet absorption band-systems of PbO, PbS, PbSe and PbTe 449
Valves and impurity semi-conductors, A theory of flicker noise in; and discussion 366, 403
Valves, thermionic, Spontaneous fluctuations of electricity in; and discussion 375, 388, 403
Vigoureux, P.: Sensitivity and impedance of electro-acoustic transducers . . 19
Viscosity and thermal conductivity of liquids, Effect of temperature on. . . 1030

Ward, F. A. B.: A mechanical model illustrating the uranium chain reaction . 113
Ward, F. A. B.: A simple optical model demonstrating the principle of the Bragg x-ray spectrometer 111
Water under hydrostatic tension, The behaviour of: III 199
Wavefront-shearing interferometer 940
Weil, R.: A note on the intermittency effect 161
Weil, R.: The variation of the reflectivity of nickel with temperature . . . 781
Wise, M. E., *see* Lord, Mary P.
Witty, R., *see* Avery, D. G.
Wolf, E. and Preddy, W. S.: On the determination of aspheric profiles . . 704
Wyllie, G.: The hole theory of diffusion 694

X-ray spectrometer, Bragg, Optical model demonstrating 111

Zinc sulphide, Short period time variation of the luminescence of; and corrigenda 473, 711

INDEX TO REVIEWS OF BOOKS

PAGE
Aharoni, J. : *Antennae : An Introduction to their Theory* 328

Cady, W. G. : *Piezoelectricity* 513
Carslaw, H. S. and Jaeger, J. C. : *Conduction of Heat in Solids* 519
Cohen, B. S. : *A Handbook of Telecommunication* 158
Cosslett, V. E. : *Introduction to Electron-Optics* 327

De Vries : *German-English Science Dictionary* 1046
Dirac, P. A. M. : *The Principles of Quantum Mechanics* 1047

Frenkel, J. : *The Kinetic Theory of Liquids* 516

Gaydon, A. G. : *Dissociation Energies and Spectra of Diatomic Molecules* . . 1044
Gill, F. C. : *The Vector Operator j* 520
Graves, L. M. : *The Theory of Functions of Real Variables* 517

Hercock, R. J. : *Photographic Recording of Cathode-ray Tube Traces* . . . 1048
Huxley, L. G. H. : *A Survey of the Principles and Practice of Waveguides* . . 1042

Jaeger, J. C. : *see* Carslaw, H. S.
Jeffreys, B. S. : *see* Jeffreys, H.
Jeffreys, H. and Jeffreys, B. S. : *Methods of Mathematical Physics* 512

Luckiesh, Matthew : *Light, Vision and Seeing* 1048

McLachlan, N. W. : *Theory and Application of Mathieu Functions* 711
Mathematical Tables Project : *Tables of Fractional Powers* 518
Mathematical Tables Project : *Tables of Spherical Bessel Functions* 520

Parke, Nathan Greer : *Guide to the Literature of Mathematics and Physics* . . 1046
Pidduck, F. B. : *see* Sas, R. K.
Pipes, L. A. : *Applied Mathematics for Engineers and Physicists* 515

Relton, F. E. : *Applied Bessel Functions* 1047
Richardson, E. G. : *Physical Science in Art and Industry* 517
Richardson, E. G. : *Sound: A Physical Textbook* 1045
Russell, B. : *Physics and Experience* 509

Sas, R. K. and Pidduck, F. B. : *The Metre-Kilogram-Second System of Electrical Units* 903
Sommer, A. : *Photoelectric Cells* 157

Tolansky, S. : *Introduction to Atomic Physics* 520

Watson, W. H. : *The Physical Principles of Wave Guide Transmission and Antenna Systems* 902
Weatherburn, C. E. : *A First Course in Mathematical Statistics* 160
Willmer, E. N. : *Retinal Structure and Colour Vision* 518
Wright, W. D. : *Researches on Normal and Defective Colour Vision* 326
Wrinch, D. : *Fourier Transforms and Structure Factors* 1044

Young, V. J. : *Understanding Microwaves* 159